PLoS one

A Deep Insight into the Sialotranscriptome of the Gulf Coast Tick, *Amblyomma maculatum*

Shahid Karim[1], **Parul Singh**[1], **José M. C. Ribeiro**[2]*

1 Department of Biological Sciences, The University of Southern Mississippi, Hattiesburg, Mississippi, United States of America, 2 Vector Biology Section, Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland, United States of America

Abstract

Background: Saliva of blood sucking arthropods contains compounds that antagonize their hosts' hemostasis, which include platelet aggregation, vasoconstriction and blood clotting; saliva of these organisms also has anti-inflammatory and immunomodullatory properties. Perhaps because hosts mount an active immune response against these compounds, the diversity of these compounds is large even among related blood sucking species. Because of these properties, saliva helps blood feeding as well as help the establishment of pathogens that can be transmitted during blood feeding.

Methodology/Principal Findings: We have obtained 1,626,969 reads by pyrosequencing a salivary gland cDNA library from adult females *Amblyomma maculatum* ticks at different times of feeding. Assembly of this data produced 72,441 sequences larger than 149 nucleotides from which 15,914 coding sequences were extracted. Of these, 5,353 had >75% coverage to their best match in the non-redundant database from the National Center for Biotechnology information, allowing for the deposition of 4,850 sequences to GenBank. The annotated data sets are available as hyperlinked spreadsheets. Putative secreted proteins were classified in 133 families, most of which have no known function.

Conclusions/Significance: This data set of proteins constitutes a mining platform for novel pharmacologically active proteins and for uncovering vaccine targets against *A. maculatum* and the diseases they carry.

Introduction

Saliva of ticks is a complex mixture of pharmacologically active compounds that interact with their host's hemostasis (the combined result of platelet aggregation, vasoconstriction and blood clotting) and inflammatory reactions that might disrupt feeding [1,2]. In addition to helping ticks to feed by its biological activities, tick saliva can also enhance pathogen transmission, either by co-feeding ticks [3], or by helping the survival of the pathogen in its adaptation to the new vertebrate host [4,5,6,7]. Because of its dual role in feeding and pathogen transmission, anti-tick saliva vaccines have been proposed as both anti-tick and/or anti tick-borne disease vaccine targets [8,9,10].

The combined mixture of these salivary compounds, the sialome (from the Greek sialo = saliva), can be partially uncovered by sialotranscriptome studies that are revealing several hundred different proteins in different tick species [2]. These studies also reveal that the salivary proteins of these organisms are at a very fast pace of evolution, probably due to their host's immune pressure, and perhaps for this reason individual genera, or subgenera of such arthropods have several unique protein families (= no similarity matches to known proteins at the amino acid

sequence level) and that many salivary proteins are products of gene duplication [11], thus creating the scenario for divergent evolution among members of these families.

Sialotranscriptomes have been produced, so far, by Sanger sequencing of cDNA libraries. The number of such sequences, also known as expressed sequence tags (EST's) varies per study from 500 to several thousands [12,13,14,15,16,17,18,19,20,21,22,23,24]. However, the upcoming of the "next generation" sequencing in the form of pyrosequencing allows for cheaper sequencing of millions, not thousands, of sequences thus permitting a much deeper insight into rarer transcripts than previously done.

The Gulf Coast tick *Amblyomma maculatum* is found in the American states surrounding the Gulf and in the Eastern Atlantic region. It is a catholic 3-host tick, immatures of which feed on small rodents and ground birds, while the adults feed on large mammals, being an economical pest of cattle [25]. It can produce tick paralysis in humans [26,27], to transmit *Hepatozoon americanum* to dogs [28,29,30] and to harbour *Rickettsia parkeri* in the US [31]. We here report an annotated catalogue of salivary gland expressed transcripts from adult females of *A. maculatum* resulting from over 1.5 million pyrosequencing sequences, representing the deepest analysis of any sialotranscriptome performed so far. This catalogue

should represent a knowledge platform for the discovery of novel pharmacologically active proteins, novel vaccine targets and novel immunoepidemiological markers of tick exposure.

Methods

Ticks and Salivary Gland (SG) Preparation

Pathogen-free *Am. maculatum* adult ticks were obtained from Oklahoma State University's tick-rearing facility. All unfed ticks were maintained in the laboratory at 23°C and >90% relative humidity under a 14-hour light/10-hour dark photoperiod before infestation on a sheep according to the methods of Patrick and Hair [32]. Adult ticks were fed on sheep in accordance with protocol # 10042001 approved by the Institutional Animal Care and Use Committee at the University of Southern Mississippi specifically for this study. The partially blood-fed (2, 3, 4, 5, 6, 7, & 9 days post attachment) female adult ticks were dissected within four hours after removal from the sheep. Tick SGs were dissected from 20–30 female ticks from each feeding stage. The dissecting solution was ice cold 100 mM 3-(N-Morpholino)-propanesulfonic acid (MOPS) buffer containing 20 mM ethylene glycol bis-(β-aminoethyl ether)-N, N, N′, N′-tetraacetic acid (EGTA), pH 6.8. After removal, glands were washed gently in the same ice-cold buffer. The dissected SGs were stored immediately after dissection in RNAlater (Ambion Inc., Austin, TX, USA) prior to isolating mRNA.

RNA Preparation

Poly A$^+$ mRNA was isolated from tick SGs using the Illustra™ QuickPrep micro mRNA purification kit (GE Healthcare, Piscataway, NJ, USA) following the manufacturer's protocol. The quality of the mRNA samples was confirmed by lab-on-chip analysis using the 2100 Bioanalyzer (Agilent Technologies, Inc., Santa Clara, CA, USA). The mRNA quantity was determined by a Nanodrop, and the mRNA samples (A260/280>1.8) were pooled for further cDNA synthesis.

Library Preparation and Sequencing

Library preparations for GS FLX titanium (Roche/454 Life Sciences, Branford, CT, USA) sequencing were developed in the Center for Genomics and Bioinformatics, Indiana University, based on methods for use in GS FLX standard sequencing described in [33] with modifications (K. Mochaitis, unpublished). Briefly, cDNA was synthesized from 1.2 μg of pooled mRNA (2–9 days-post feeding) in a manner similar to Clontech SMART protocols, using primers optimized for the 454 sequencing process and amplified by PCR to generate dsDNA. The cDNA library was normalized using Trimmer cDNA normalization kit (Evrogen, Moscow, Russia) according to manufacturer's instructions. Normalized DNA was then fragmented by sonication, and ends were enzymatically blunted and ligated to customized 454 adaptors. Amplification of ligation products exploited adaptor-mediated PCR suppression [33,34]. This procedure induces homo-mediated fragment hairpins, thereby severely limiting amplification of mis-ligated products. All amplification steps utilized high-fidelity polymerases. Final library was size selected by excision of the 500–800-bp fraction from the agarose gel. This size selection may have reduced the probability of finding small transcripts such as single Kunitz domain proteins and antimicrobial peptides. Emulsion PCR reactions were performed according to the manufacturer (Roche 454 Life Sciences). To optimize the pyrosequencing throughput, the final libraries were titrated by emulsion PCR bead enrichment prior to sequencing. Sequencing of the salivary cDNA library was performed on a picotitre plate according to the manufacturer's instructions, and yielded 560.4 Mb of sequence data in 1,626,969 read with an average of 344 nucleotides (nt) in length. Sequencing adapters (A and B) were automatically removed from the reads using signal processing software (Roche 454 Life Sciences).

Bioinformatics Tools Used

The blastn tool (performed locally from executables obtained at the NCBI FTP site ftp://ftp.ncbi.nih.gov/blast/executables/) [35] and CAP3 assembler [36] were used for EST clusterization, by a decreasing word size inclusion strategy. A master program sequentially sent each EST to be blasted against all ESTs using an initial word size of 200 (BLASTN switch -W 200) and a maximum limit of 1,000 matches (-v 1000, using tabular output mode -m 8). Matches were marked as collected as they were retrieved from the blastn program, and these matched sequences were not sent for blast-n when their turn arrived, thus avoiding duplicating the BLAST task. All matches were collected into FASTA and qual-formatted files and fed as input to the CAP3 assembler. The CAP3-outputted FASTA of the assembled data was obtained, including quality files that were then the starting point for the next cycle. This second iteration was done with a word size of 134, the output of which was in turn used for the next assembly round, but now using a word size of 90, then 60, then two more rounds of 40 to produce the final assembly shown in Additional file S1. This assembly strategy is easy to parallelize, allowing for large data sets to be clusterized. In the current case, a total of 48 CPUs were used for 24 hours. The final assembly output was piped into a tab-delimited file that was imported into an Excel spreadsheet, which includes for each assembled contig the number of reads and the list of unique names for each read, to facilitate counting the contribution of different libraries for the final assembly. These operations were automated by a program written in Visual Basic (VB) named Megacluster and associated blaster clients.

Table 1. Nature and abundance of reads for extracted coding sequences (CDS) from the sialotranscriptome of *Amblyomma maculatum*.

Class	Number of CDS	Number of Reads	Reads/CDS	Percent of Total Reads
Secreted	3475	296284	85.3	23.7
Housekeeping	7856	787547	100.2	62.9
Unknown	4167	160529	38.5	12.8
Transposable Elements	316	7577	24.0	0.6
Total	15814	1251937		

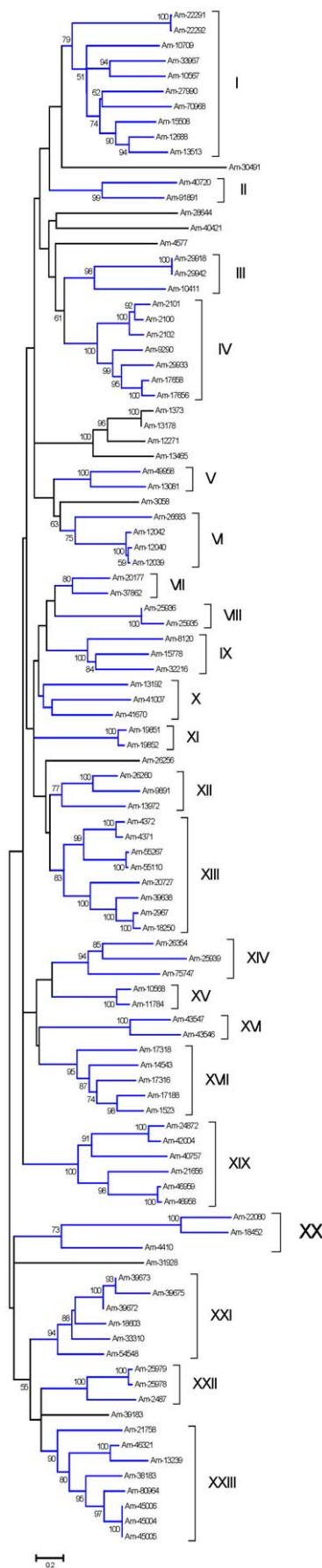doi:10.1371/journal.pone.0028525.t001

Figure 1. The salivary Kunitz family of proteins in *Amblyomma maculatum*. Bootstrapped phylogram (1,000 iterations) resulting from the alignment of 101 full-length protein sequences containing one or more Kunitz domains. The numbers on the nodes indicate the bootstrap support, and the bar at the bottom indicates 20% amino acid divergence. Clades with more than 70% bootstrap support are indicated by Roman numerals.
doi:10.1371/journal.pone.0028525.g001

Segments of the six-frame translations of the contigs starting with a methionine found in the first 300 predicted amino acids, or the predicted protein translation in the case of complete coding sequences, were submitted to the SignalP server [37] to help identify translation products that could be secreted. To obtain insight on the nature of the transcripts, blastx, blastn, or rpsblast searches of the contigs against several databases were performed. These databases were: the non redundant protein database (NR) of the National Center for Biotechnology Information (NCBI); the Swissprot database; the gene ontology(GO) FASTA subset [38]; the tick salivary sequences described in a previous review [2]; custom downloaded databases from GenBank containing mito-chondrial and rRNA nucleotide sequences; and the conserved domains database of NCBI [39] containing the KOG [40], PFAM [41], and SMART [42] motifs.

Sequences matching 50% or more of the length of proteins in the NR, Swissprot, or tick salivary databases had their coding sequences (CDS) automatically extracted by another program written in VB by the senior author to compose the Spreadsheet S2. Because pyrosequencing introduces insertion/deletion (indel) errors in the sequences, this program also recognizes frame shifts in the blastx-derived alignments and marks as N (instead of A, T, C, or G) the nucleotides in the region of frame shift and either subtracts or adds one N to correct the alignment. Spreadsheet S2 was compared by blastp and rpsblast.

Deducted protein sequences were also sent to the SignalP server, to the TMHMM server [43] to detect membrane helices, the NetOglyc server to detect possible mucin-type galactosylations [44] and to the ProP server [45] to identify putative furin-processed protein cleavage sites. The protein sequences in Spreadsheet S2 were also clusterized progressively from 25% similarity to 99% similarity over 50% of the length of the larger sequence, thus helping to identify related protein families.

To functionally classify the protein sequences, another VB program was written that took into consideration key words in the BLAST matches of the Swissprot, GO, tick subset, KOG, PFAM, and SMART databases, as well as their e values, plus the results for SignalP, transmembrane domains, and glycosylation (not used in Spreadsheet S1) to produce nearly 30 functional categories, as indicated in the Spreadsheets S1 and S2. The final results presented were in many cases manually corrected.

Phylogenetic analysis and statistical neighbor-joining bootstrap tests of the phylogenies were done with the Mega package [46] after sequence alignment performed by ClustalW [47].

Results and Discussion

A total of 1,626,969 pyrosequencing reads were assembled into 190,646 contigs, including singletons. A subset of this data containing only sequences larger than 149 nt yielded 72,441 sequences; these are displayed in Spreadsheet S1. This subset of 72,441 contigs contained 1,498,171 reads, or 92% of the totality of reads. Other parameters of the original reads and assemblies are provided in Spreadsheet S1. A program written by JMCR in VB (see Methods) extracted the coding sequences from the assembled data having at least five reads per contig, generating 15,814 CDS,
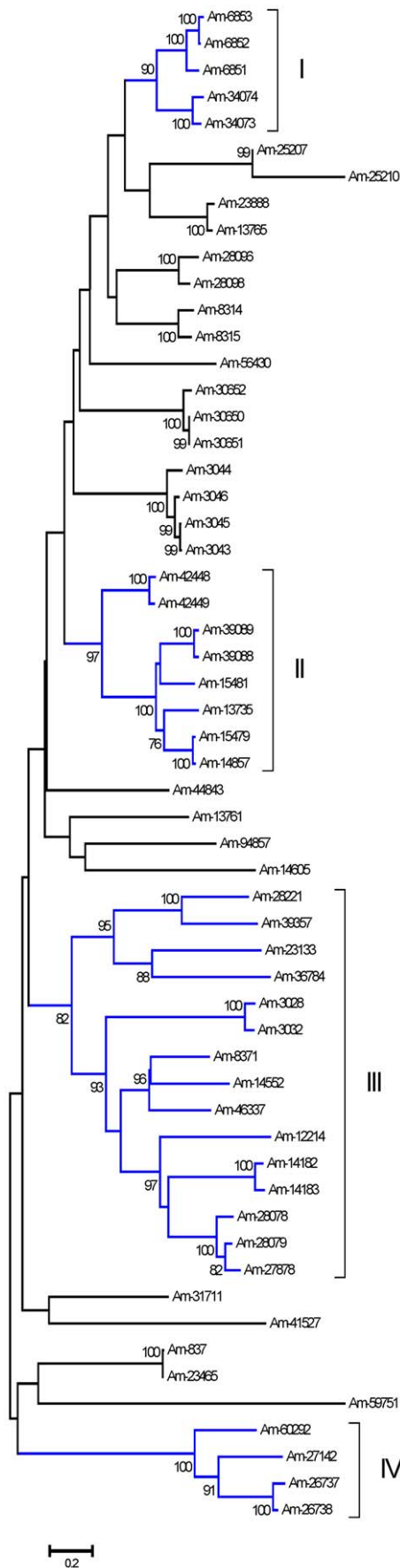
Figure 2. The salivary TIL domain family of proteins in *Amblyomma maculatum*. Bootstrapped phylogram (1,000 iterations) resulting from the alignment of 57 full-length protein sequences containing the Kunitz domain. The numbers on the nodes indicate the bootstrap support, and the bar at the bottom indicates 20% amino acid divergence. Clades with more than 80% bootstrap support are indicated by Roman numerals.
doi:10.1371/journal.pone.0028525.g002

which derived from 1,251,937 reads, or from 77% of the totality of reads (Table 1). Of these, 5,353 had >75% coverage to their best match in the NR database, and 8,785 had e values by blastp ≤1E-15 when compared with the same database. This annotated set is provided in Spreadsheet S2. Four broad categories of expressed genes are observed in Spreadsheet S2 and summarized in Table 1. The putatively secreted (S) category contained 24% of the reads and had on average 85 sequences per CDS; the housekeeping (H) category had 63% of the reads with an average of 100 reads/CDS; while 13% of the reads with an average of 39 reads/CDS were not classifiable, constituting the Unknown (U) group. Sequences deriving from transposable elements (TE) accounted for 316 contigs, with an average of 24 reads/CDS, and representing 0.6% of the reads. TE-related sequences may indicate either the presence of active transposition in the tick or, more likely, the expression of sequences suppressing transposition. Low-level expression of TE sequences has been a relatively common finding in previous sialotranscriptomes.

The following text is a guide to browsing Spreadsheet S2.

Housekeeping (H) Genes

The 7,857 CDS attributed to H-class genes expressed in the SGs of *Am. maculatum* were further characterized into 23 subgroups according to function (Table S1 and Spreadsheet S2). Normalization of the library plus the sheer number of reads allowed an unusually deep recovery of transcripts not normally found with limited Sanger-based transcriptomes, such as transcription factors. For example, the *Am. maculatum* homolog of the Enhancer-of-yellow-2 transcription factor was assembled with 920 reads, as was the NFAT transcription factor, assembled from 770 reads. Spreadsheet S2 presents a total of 162 putative transcription factor coding sequences. Of importance for salivary function, 395 coding sequences associated with protein export machinery were retrieved. The vacuolar sorting protein VPS28 was assembled from 1,333 reads. Many vesicular transport-associated proteins—as well as SNARE proteins and members of the signal peptidase complex—were identified and annotated. Native immunity components include proteins annotated as Toll-like receptors and thio-ester/complement-like proteins. Detoxification enzymes were also found, such as sulfotransferases that might be associated with dopamine detoxification [48,49], a main secretagogue for tick SGs [50,51]. Enzymes dealing with oxidative detoxification are also abundant, some of which, such as selenoproteins, peroxidasins (haem-peroxidases) and superoxide dismutase may actually be secreted and antagonize inflammatory host responses containing superoxide, nitric oxide or peroxinitrite [52,53]. Enzymes of the cytochrome P450 family also abound in the sialotranscriptome where 65 contigs were assembled, several of which appear as full length proteins, such as Am-35392 and Am-38412.

Transposable Elements (TE)

Both class I (retrovirus-like elements) and class II (coding solely for a transposase protein, and having inverted terminal repeats) transposon coding sequences were found in the sialotranscriptome of *Am. maculatum*. All class I–derived coding sequences were

truncated by having premature stop codons when compared with active transposons. These products probably act as regulators of transposition or represent remnants of previously active transposition events in the tick genome. On the other hand, seemingly complete transposases of class II elements were identified for the PIF, POGO, and TC1 families, suggesting active transposition of these elements in the tick genome or relatively recent activity.

Possibly Secreted (S) Class of Expressed Genes

A total of 3,475 coding sequences, derived from 296,284 reads, are possibly associated with a bloodfeeding role as secreted salivary proteins (Table 1 and Spreadsheet S2). These include previously known gene families [2] such as diverse classes of protease inhibitors, metalloproteases, immunomodulators, antimicrobial peptides, basic tail, and glycine-rich peptides. Several protein families were discovered to be specific to *Amblyomma* ticks (because they provide no similarities to other known proteins by blastp), many of which are found to date only on *Am. maculatum*. A total of 133 distinct family classes are annotated in Table S1 and Spreadsheet S2. These families are organized in seven general classes, as described below:

Protease inhibitor-containing domains

A total of 379 CDS representing 2.61% of the reads code for proteins containing signatures of proteins previously associated with a protease inhibitory function, which are either ubiquitous or particular to ticks. A more detailed analysis of these transcripts follows.

Kunitz domain-containing proteins. Kunitz domain-containing proteins abound in tick sialotranscriptomes [2] as well as in those of the hematophagous flies of the genera *Culicoides* [54,55] and

Simulium [56], but not mosquitoes or sand flies. Proteins containing single or multiple Kunitz domains were described and functionally characterized in ticks, such as Ixolaris, a double-Kunitz protein, and Pentalaris, containing five domains [57,58,59,60,61], both acting as blood-clotting inhibitors of the extrinsic pathway. The Kunitz fold can also perform functions beyond protease inhibition, such as ion channel inhibition [62,63,64,65]; in *R. appendiculatus*, a modified Kunitz domain peptide [66] was shown to activate maxiK channels in an *in vitro* system, suggesting a vasodilator function.

From the sialotranscriptome of *Am. maculatum*, a total of 215 CDS, extracted from 18,071 reads, were assigned to the Kunitz family of proteins based on their SMART signature. Within this group, there are proteins containing from one to six Kunitz domains. From this set, 101 are complete from starting methionine to stop codon and contain a signal peptide indicative of secretion. Several of these proteins contain one or more proP signals indicative of furin processing [45], and thus could be pro-proteins. ClustalW alignment and neighbor-joining phylogenetic analysis of this set (Figure 1) shows the diversity of this class of proteins. Gene duplication has been proposed before for explaining the expansion of protein families associated with blood feeding [11,67]; the presence of 23 clades (defined by related proteins belonging to a clade with >70% bootstrap support) containing these 101 proteins are indicative of this process. Notice that the tree includes proteins that are very similar to each other, such as the two top sequences, or bottom three sequences, in Figure 1. These sequences could represent alleles from the same gene. It is possible that members of the same clade share the same function and are expressed at different times during feeding to avoid their host immune system. This expansion of the Kunitz family of proteins has been observed before in other tick sialotranscriptomes, but never to this degree.
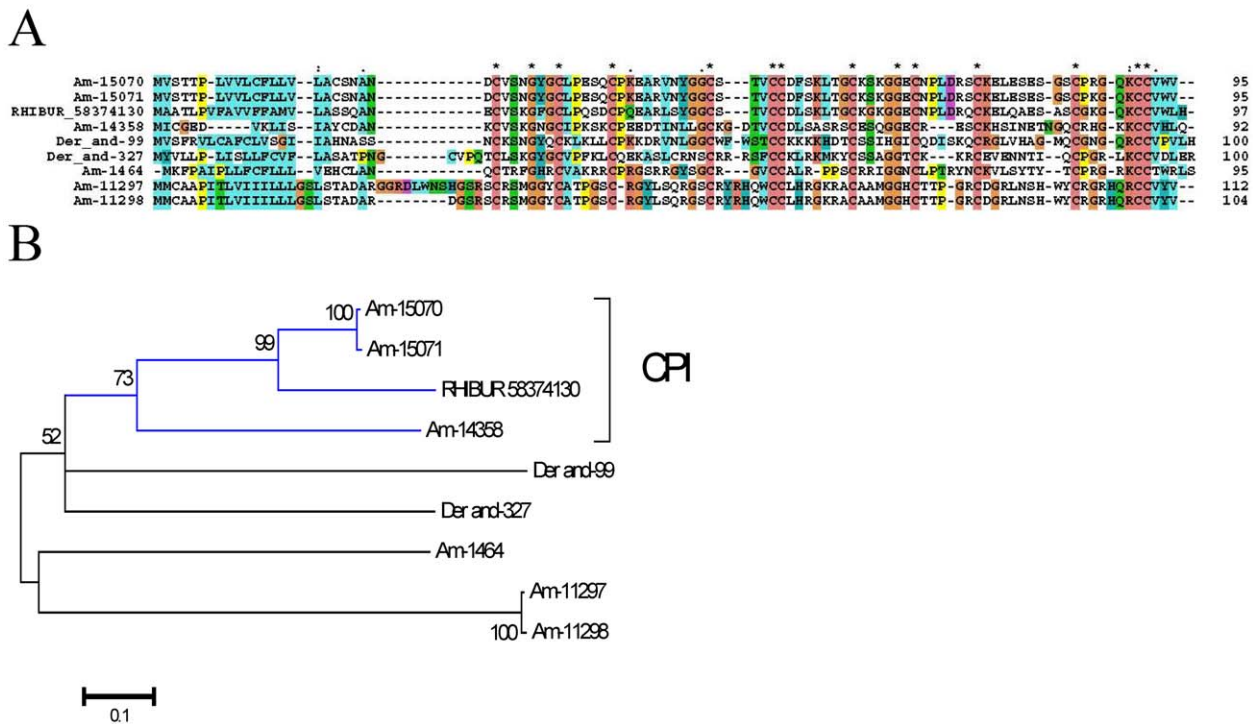

Figure 3. The tick salivary carboxypeptidase inhibitor family of proteins. (A) Alignment of *Amblyomma maculatum*, *Dermacentor andersoni*, and *Rhipicephalus bursa* proteins. **(B)** Bootstrapped phylogram (1,000 iterations) resulting from the alignment in **(A)**. The numbers on the nodes indicate the bootstrap support above 50%. The clade indicated by CPI contains the functionally characterized *R. bursa* protein. The bar at the bottom indicates 10% amino acid divergence. The *Dermacentor* proteins were deducted from ESTs deposited in DBEST.
doi:10.1371/journal.pone.0028525.g003

TIL domain-containing proteins. The canonical TIL (for trypsin inhibitor-like) domain contains ten cysteines forming five disulphide bonds and is found in many protease inhibitors. These polypeptides may also exert antimicrobial function [68]. Members of this family have been found ubiquitously in bloodfeeding insect and tick sialomes, but very few have been characterized. A tick hemolymph antimicrobial peptide (AMP) was previously reported to be a member of this family [69]. More recently, tick proteins containing TIL domains were characterized from *R. microplus* as subtilisin inhibitors with antimicrobial activity and expressed in various tick organs, including the SGs [70].

Eighty-five CDS were extracted from 7988 reads originating from the sialotranscriptome of Am. maculatum, 57 of which are complete from Met to stop codon and have a predicted signal peptide indicative of secretion. These proteins can contain from one to four TIL domains. Some have no TIL domain but exhibit similarities to proteins containing the domain and for this reason are included in this group. Phylogenetic analysis of the 57 related sequences (Figure 2) shows four relatively large clades with strong bootstrap support. As in the case of the Kunitz domain-containing proteins, these different TIL domain-containing proteins could have appeared as products of gene duplication, and clade members may have the same function but different antigenicities.

Thyropins. Thyropins are motifs found in thyroglobulins and in cysteine protease inhibitors of the actiniam-derived equistatin protein [71,72,73]. Equistatin has three thyropin domains, two of which were shown to be involved in protease inhibition.[73]. They are recognizable by the SMART TY domain match. A protein deducted from the sialotranscriptome of *Amblyomma variegatum* was found to have two thyropin domains [15], and one-domain proteins were found in the sialome of *R. sanguineus* [20]. Six coding sequences containing TY domains were assembled from 1,917 reads, all of which have two TY domains. Am-4121 was assembled with 1,329 reads, being the most abundantly expressed of this group, more so considering the library was normalized. All these proteins provide best matches to a *R. sanguineus* salivary thyropin deposited in the NR database, producing 54–66% identity at the amino acid sequence level. Interestingly, Am-4121 has a serine-rich carboxyterminus with 26 potential galactosylation sites, indicating it has thyropins and mucin domains. Proteins deducted from ESTs of *R. microplus* and *Am. variegatum* also have these mucin domains, and these can be inspected in the BLAST matches to the tick-tb2 database on Spreadsheet S2.

Cystatins. Cystatins are cysteine protease inhibitors of nearly 100 amino acids (aa) in length. Cystatins have been previously found in both hard and soft tick sialotranscriptomes. Two salivary cystatins from *I. scapularis* have been functionally characterized as inhibitors of cathepsins L and S, to inhibit inflammation, suppress dendritic cell maturation, and serve as vaccine targets [74,75]. Twenty-five CDS were assembled from 1,100 reads originating from the *Am. maculatum* sialotranscriptome; 15 of these 25 appear full length and have a signal peptide indicative of secretion.

Serpins. Serpins are a ubiquitous protein family associated with the function of serine protease inhibition, from which the family name derives. A single tick salivary serpin from *I. ricinus* has been shown to inhibit vertebrate elastase and to have immunosuppressive activity. [76,77] Another salivary serpin from the same tick inhibits cathepsin G and chymase [78]. Tick serpins have been proposed as anti-tick salivary vaccines, including non-salivary expressed serpins. [79,80] The sialotranscriptome of *Am. maculatum* reveals 32 CDS for members of the serpin family, assembled from 1,100 reads. Eight of these CDS appear as full length and with a signal peptide indicative of secretion.
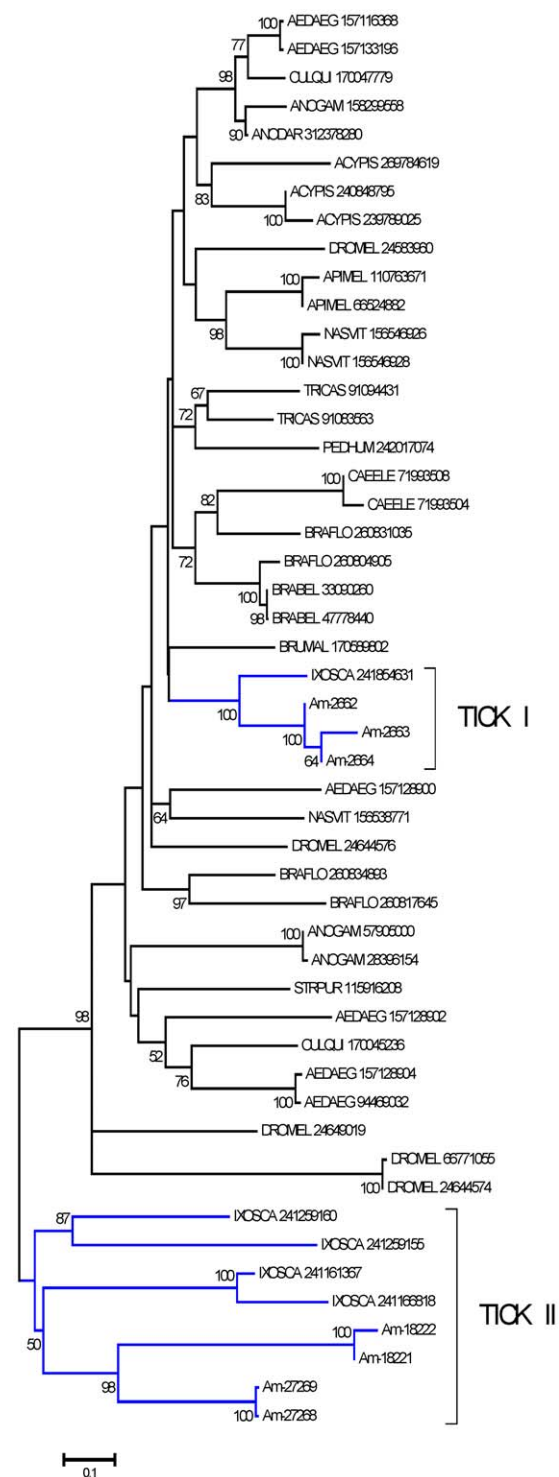

Figure 4. Phylogram of the phosphatidylethanolamine binding protein (PEBP) family of selected arthropods. The bootstrapped phylogram (1,000 iterations) was obtained from the alignment of deducted *Amblyomma maculatum* proteins with homologs found in the non-redundant protein database of the NCBI. The *Am. maculatum* protein names start with Am- and the remaining proteins are named by the first three letters of the genus name followed by the first three letters of the species name followed by their NCBI gi| accession number. The number at the nodes indicates the bootstrap support, and the bar at the bottom indicates 10% amino acid divergence. Two tick clades are indicated by Roman numerals.
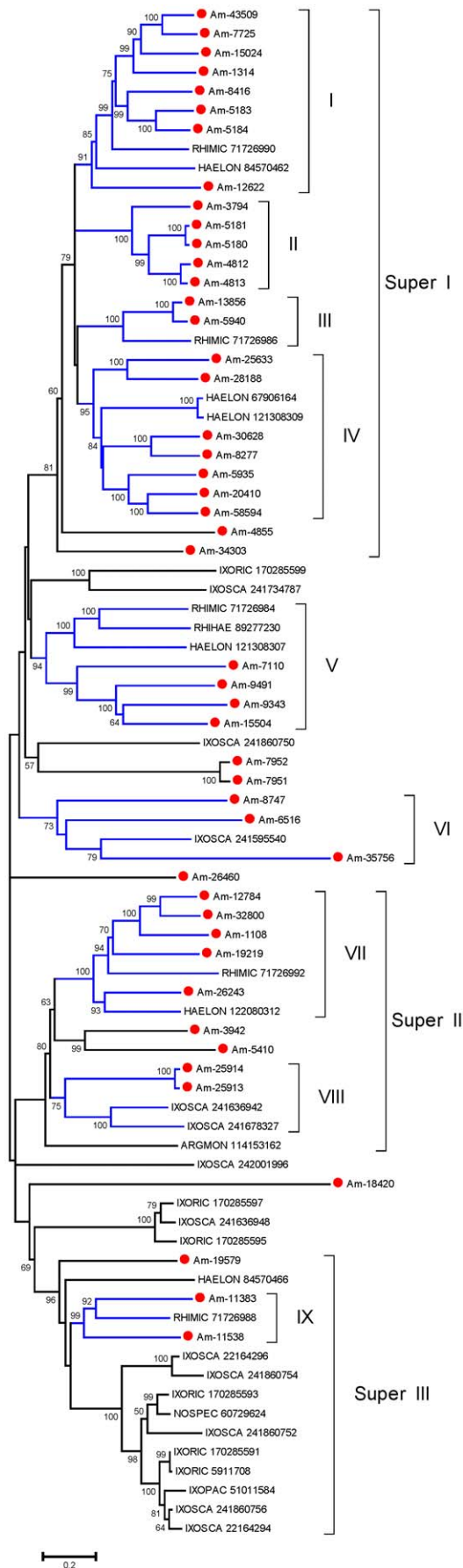doi:10.1371/journal.pone.0028525.g004

Figure 5. Phylogram of the tick salivary metalloproteases of the reprolysin family. The bootstrapped phylogram (1,000 iterations) was obtained from the alignment of deducted *Amblyomma maculatum* proteins with homologs found in the non-redundant protein database of the NCBI. The *Am. maculatum* protein names start with Am- and are recognized by a red circle marker; the remaining proteins are named by the first three letters of the genus name followed by the first three letters of the species name followed by their NCBI gi| accession number. The number at the nodes indicates the bootstrap support, and the bar at the bottom indicates 20% amino acid divergence. Clades and superclades with strong bootstrap support are indicated with Roman numerals.
doi:10.1371/journal.pone.0028525.g005

Kazal domain. The Kazal domain is also associated with serine protease inhibitors and antimicrobial activity [81,82]. Members of this family have been found in anopheline and culicine mosquito sialotranscriptomes and also occasionally in tick sialotranscriptomes. In the mosquito *Aedes aegypti*, a salivary-expressed Kazal-containing peptide was shown to inhibit thrombin and plasmin with high affinity [83]. The Kazal domain can also be found in multidomain proteins, as was the case of a protein found in the sialotranscriptome of *Am. variegatum* containing an insulin growth factor binding domain in tandem with a Kazal domain, suggesting this protein could function in the modulation of signal transduction cascades.

The CDS Am-25261 is such a complex protein, containing a Kazal domain in its amino terminal region (aa 57–101), an EF-hand domain (aa 150–222), and a von Willebrand factor type C domain in the carboxyterminus (aa 232–277). This protein is 82% identical in amino acid sequence to the follistatin-related protein (FRP) of the tick *Haemaphysalis longicornis* [84]. Transcripts for this protein in *Haemaphysalis* were expressed in several organs and may indicate it has a housekeeping function; however, vertebrate FRP—which are 40% similar to the tick protein—can modulate ion transport in neurons in a way that makes them hyperpolarized and with higher excitation thresholds [85].

Carboxypeptidase inhibitors (CI). This is a protein family found to date only in metastriate ticks first identified in *R. bursa*, from where the proteins functionally characterized and crystallized [86,87]. CI homologues from *Ha. longicornis* were also characterized [88]. These inhibitors may affect fibrinolysis by inhibiting plasma carboxypeptidase B, also known as thrombin-activatable fibrinolysis inhibitor [86]. Six CDS coding for members of this family were deducted from 425 reads derived from the sialotranscriptome of *Am. maculatum*. Alignment of these sequences with the canonical *R. bursa* protein plus *Dermacentor andersoni* sequences obtained from a previous review [2] shows 12 conserved cysteines, 4 of which are in 4 cysteine knot (CC) configuration (Figure 3A). Only three of the *Am. maculatum* proteins are within the same clade as the *R. bursa* protein (Figure 3B), indicating the diversity of this family.

Phosphatidylethanolamine-binding protein family. This is a ubiquitous protein family that has been associated with serine protease inhibition [89,90], although such activity has never been functionally characterized from any bloodsucking arthropod to date. Seven CDS were extracted from 815 reads deriving from the *Am. maculatum* sialotranscriptome. Overall, these proteins share less than 40% similarity and represent at least three expressed genes with possible alleles, as indicated by phylogenetic analysis of their closest invertebrate proteins (Figure 4), where the robust clade named TICK I has at least one gene, and the clade marked TICK II has at least 2 genes.

Enzymes

A total of 415 CDS, representing 4% of the reads, code for proteins containing enzyme signatures and a secretion signal, or

are related to secreted enzymes. Some of these enzymes could actually be destined to the ER, the Golgi apparatus, or to lysosomes and not actually secreted.

Proteases. Proteases represent the bulk of the extracted sequences coding for enzymes, represented by 311 CDS deriving from 3.2% of the reads. Among this class of enzymes, metalloproteases predominate, with 268 extracted CDS carrying 2.6% of the reads. Metalloproteases of the reprolysin family have been commonly found in tick sialotranscriptomes [2,91,92]. These enzymes can be identified by the CDD domain CDD|58573 cd04272, ZnMc_salivary_gland_MPs, which is quite specific for tick salivary reprolysin metalloproteases. In *I. scapularis*, metalloproteases of this type were associated with fibrinogenolytic and anti-angiogenic activities found in this tick saliva [93,94]. *Am. maculatum* metalloproteases appear to be a highly expanded gene family, as deduced from the phylogenetic analysis of the proteins recovered with >500 aa aligned with their best matches from the NR database (Figure 5), these matches being all from ticks. Notice that there are many polyspecific clades containing *Am. maculatum* proteins in at least three super clades. This suggests the ancestral tick to prostriates and metastriates already contained at least three metalloprotease-coding genes before evolution of the ticks, with posterior gene duplication events to create the internal clades.

Metalloproteases containing the PFAM peptidase_M13_N motif include ubiquitous metalloproteases involved in peptide processing, such as the neprilysins. Neprilysins, however, are typically extracellular membrane-bound proteins of type II characterized by an uncleaved hydrophobic segment near the NH2 terminus, creating a membrane anchor [95,96]. To the extent that tick proteins have this domain, they should not appear in tick saliva unless the tick enzymes lost their membrane-

anchoring domains. The sialotranscriptome of *Am. maculatum* reveals 41 CDS producing matches to this class of enzymes, 3 of which appear to be full length (with more than 700 aa residues) and to contain a signal peptide indicative of secretion and no membrane anchor as indicated by the SignalP server [37]. For example, the tick protein coded by Am-19220 has better identity to its chicken homolog than to its *I. scapularis* homolog. While Am-19220 has a clear signal peptide indicative of secretion, the bird enzyme has a typical membrane anchor as predicted by the SignalP-HMM server [97]. Am-19220 also has, uniquely, a Kunitz domain in its amino-terminus. Am-26776 and Am-23443 are two other *Am. maculatum* neprilysins-type metalloproteases with clear signal peptide indicative of secretion. These enzymes could play a role in degrading host inflammatory peptides.

The PFAM peptidase_M2 domain identifies dipeptidyl carboxypeptidases, including the family prototype, the vertebrate angiotensin-converting enzyme. Enzymes of this family have been implicated in bradykinin degradation by the saliva of *I. scapularis* [98]. Three CDS coding for members of this family are identified in the sialotranscriptome of *Am. maculatum*, two of which appear to be full length and contain signal peptides indicative of secretion. Am-930 appears to be the most expressed, the CDS of which was assembled with 3,724 reads, more than double the reads of all neprilysins-coding CDS.

Serine proteases are commonly found expressed in arthropod sialotranscriptomes but rarely characterized for their substrate specificity. In tabanids, one such salivary enzyme was shown to have fibrinolytic activity [99]. Several serine proteases were recognized in the sialotranscriptome of *Am. maculatum*, from where 18 CDS were assembled from 2,657 reads. Six of these enzymes appear full length and with signal peptide indicative of secretion.
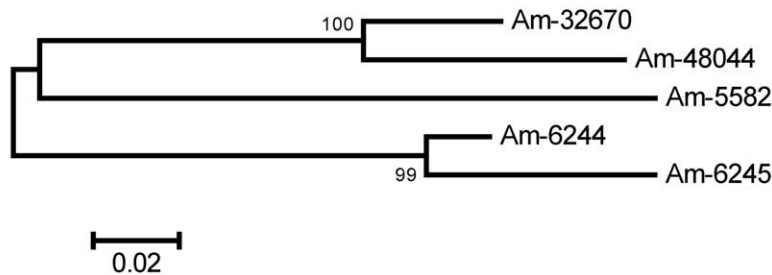

Figure 6. The salivary disintegrins 40–270 family of *Amblyomma maculatum*. (**A**) Alignment indicating conserved cysteines in black background, RGD motif in blue background, and blue bar above the sequences indicating the signal peptide indicative of secretion. The symbols at the bottom indicate (*) identity of residues, (:) conserved, and (.) less conserved residues. (**B**) Bootstrapped phylogram of the alignment in (**A**). The numbers at the nodes indicate the bootstrap support, and the bar at the bottom indicates 2% amino acid divergence.
doi:10.1371/journal.pone.0028525.g006

Proteases of the legumain family (asparaginyl endopeptidases) belong to the CD clan, family C13 of cysteine proteases, and are commonly found in tick midgut transcriptomes [100,101,102, 103,104,105,106]. Five CDS coding for legumains were identified in the present work, three of which appear to be full length. Despite having a signal peptide indicative of secretion, these proteins could be membrane bound or directed to the lysosome. Similarly, two CDS coding for cathepsin-L type of peptidases and containing a signal peptide indicative of secretion could be destined to the lysosome.

Endonucleases. These enzymes cleave RNA or DNA and have been found previously in transcriptomes of mosquitoes, sand flies [107], and ticks [2,15], but only in the mosquito *Culex quinquefasciatus* was an endonuclease shown to be secreted in saliva, and the recombinant enzyme revealed its specificity for double-stranded DNA [108]. This enzyme may help reduce the viscosity of the lacerated skin matrix but may also affect neutrophil extracellular trap formation, which is DNA rich [109]. The sialotranscriptome of *Am. maculatum* reveals several CDS coding for endonucleases, including four that appear to be full length and contain a signal peptide indicative of secretion (Am-4803, Am-12130, and Am-1109).

5′-nucleotidase/apyrases. Apyrases are commonly found in saliva of bloodsucking arthropods, where they hydrolyze ATP and ADP to AMP, thus serving an antihemostatic and antiinflammatory function, because ATP and ADP—released by injuried cells,

activated platelets, and neutrophils—are agonists of inflammation and platelet aggregation [2]. Mosquitoes and triatomines of the Triatoma genus have apyrases belonging to the 5′ nucleotidase family [110,111,112], as have soft ticks [113]. While canonical 5′ nucleotidases are extracellular enzymes bound to the cell via a glycosylphosphatidylinositol anchor, salivary apyrases lack the anchor and thus appear free in saliva [112,114,115].The sialotranscriptome of *Am. maculatum* reveals several members of the 5′ nucleotidase family including two relatively that are highly expressed (Am-4536 with 495 reads, and Am-9644 with 208 reads), both full length and not containing a glycosylphosphatidylinositol anchor as indicated by the Frag-anchor site (http://navet.ics.hawaii. edu/~fraganchor/NNHMM/NNHMM.html) [116]. Interestingly, these abundantly expressed apyrases share only 46% identities at the amino acid level and possibly represent a mechanism of antigenic variation.

Lipases and esterases. Abundant expression of mRNA coding for members of these families are found in the sialome of *Am. maculatum*, including secreted phospholipase A_2 (Am-9375 with 636 reads), a glycosylphosphatidylinositol-specific phospholipase (Am-14132 with 592 reads), and a sphyngomyelinase (Am-21256 with 527 reads). These enzymes may affect host signaling pathways in inflammation and immunity.

Glycosidases. Several contigs coding for glycosidases with signal peptide indicative of secretion were found in the
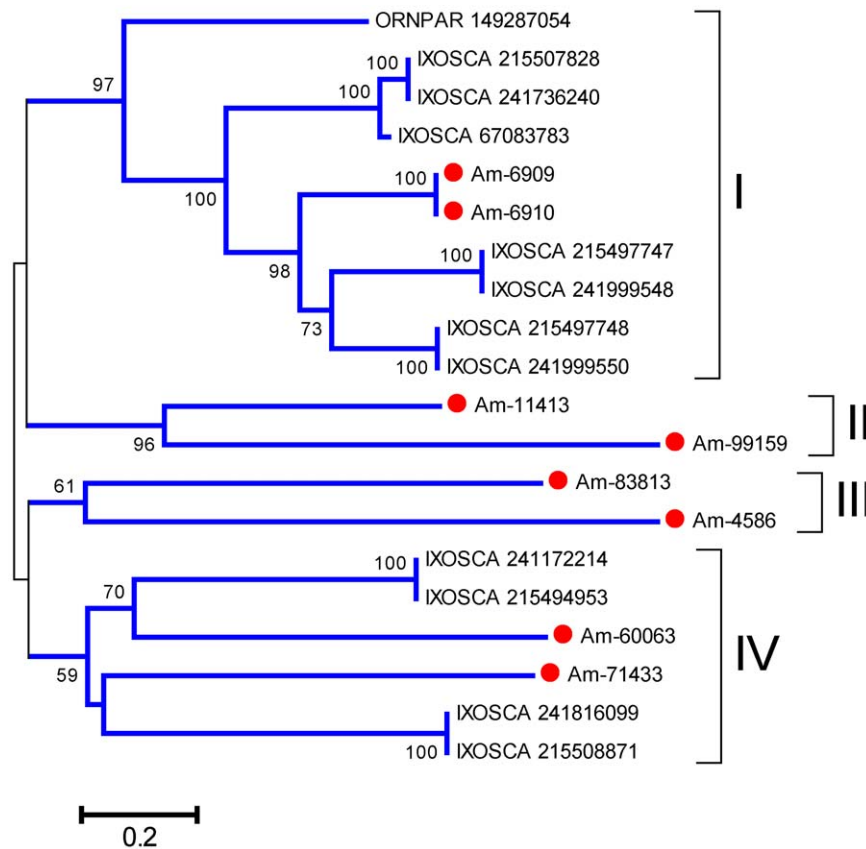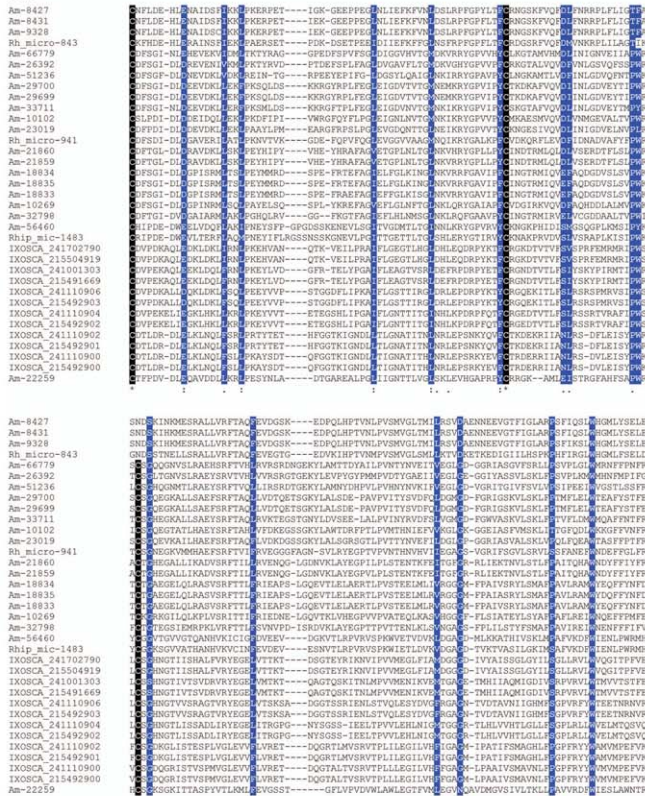

Figure 7. Phylogram of tick proteins from the 23-kDa family. The bootstrapped phylogram (1,000 iterations) was obtained from the alignment of deducted *Amblyomma maculatum* proteins with homologs found in the non-redundant protein database of the NCBI. The *Am. maculatum* protein names start with Am- and are recognized by a red circle marker; the remaining proteins are named by the first three letters of the genus name followed by the first three letters of the species name followed by their NCBI gi| accession number. The number at the nodes indicates the bootstrap support above 50%, and the bar at the bottom indicates 20% amino acid divergence. Clades and superclades with strong bootstrap support are indicated with Roman numerals.
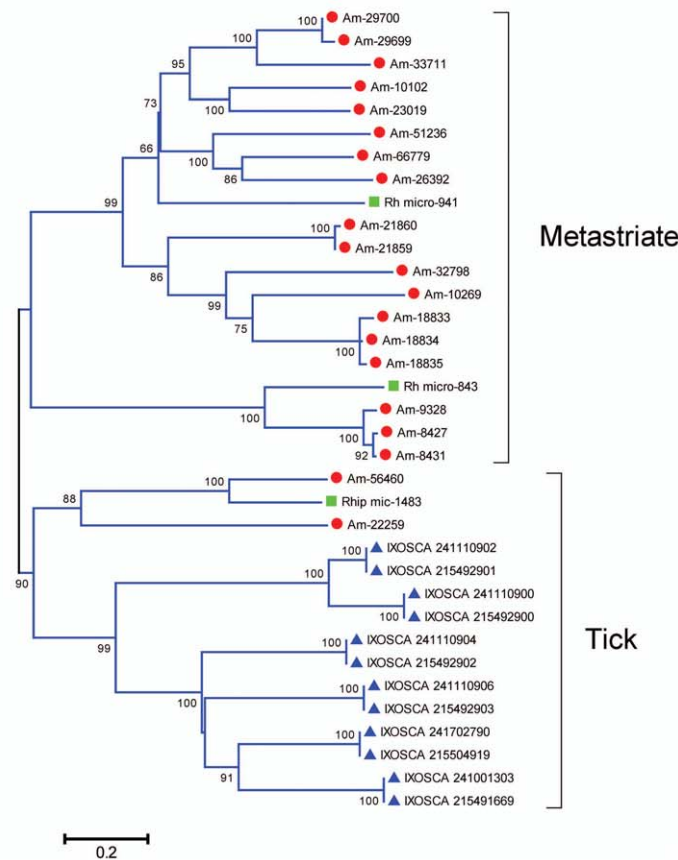doi:10.1371/journal.pone.0028525.g007

Figure 8. The one-of-each family of ticks. (A) Alignment indicating conserved cysteines in black background, conserved residues in blue background. Sequences starting with Am are from *A. maculatum* and are recognized by a red circle, those starting with IXOSCA are from *Ixodes scapularis* retrieved from GenBank and are recognized by a blue triangle, those from *Rhipicephalus* were assembled from DBEST ESTs as described before and are recognized by a green square [1]. The symbols at the bottom indicate (*) identity of residues, (:) conserved and (.) less conserved residues. **(B)** Bootstrapped phylogram of the alignment in **(A)**. The numbers at the nodes indicate the bootstrap support and the bar at the bottom indicates 20% amino acid divergence.
doi:10.1371/journal.pone.0028525.g008

sialotranscriptome, but they are most probably lysosomal enzymes and are not particularly highly expressed, with less than 40 reads on the most abundant contigs. We call particular attention to Am-23781, which is a truncated contig having the hyaluronoglucosaminidase KOG domain. Hyaluronidases have been described in hematophagous flies [99,117,118,119] and in the saliva of *Amblyomma hebraeum* [120], where they may serve to break down the host skin matrix and also affect chemokine signaling, which needs the negatively charged matrix of sulfated glycan acid for its function [121,122].

Sulfatases. Two sulfatases with signal peptide indicative of secretion were found in the sialotranscriptome. These enzymes are normally lysosomal but, if secreted, they could have a role in degrading skin mucopolysacharides, as indicated in the previous section.

Lipocalins

The lipocalin family is widespread in nature and is abundantly expressed in ticks and triatomine sialotranscriptomes as members of a large gene family. In *Am. maculatum*, we were able to extract a total of 584 coding sequences from this family with 5 or more reads, 55 of which have more than 300 reads, indicating their relatively abundant expression. Lipocalins may function as anticomplement [123], but more widely as scavengers or kratagonists [124] of biogenic amines and arachiconid acid-derived agonists of hemostasis and inflammation such as thromboxane A$_2$ and leukotrienes [125,126,127,128,129,130]. Seven of the *Am. maculatum* CDS coding for lipocalins were assembled from more than 1000 reads. These abundant lipocalins may be kratagonists of serotonin or histamine, because these agonists accumulate to near micromolar amounts during inflammation and hemostasis in contrast with nanomolar ranges of concentration for leukotrienes. It is quite remarkable that these lipocalins from *Am. maculatum* are at best only 40% identical to other tick lipocalins, suggesting a fast evolution of this protein family.

Antigen 5/CRISP/CAP

This protein family is ubiquitously found in plants and animals, having being found in the venom of vespids (whence the name antigen 5 comes [131]); in snake venoms, where they are known as cysteine rich secretory protein family (thus the name CRISP) and have toxic properties [132,133,134]; and in plants, where they are associated with pathogen responses. The superfamily CAP recovers the families known as CRISP, antigen 5, and pathogen-associated proteins from plants [135]. Virtually all sialotranscriptomes of bloodsucking arthropods have members of this family [2,124]. In stable flies, a member of the family binds immunoglobulins and may inhibit the classical pathway of complement [136,137]; an antigen 5 protein from tabanids uniquely inhibits platelet aggregation and angiogenesis [99,138,139] through acquisition of an RGD (disintegrin) domain [140]. The sialotranscriptome of *Am. maculatum* revealed several members of this family. Spreadsheet S2 presents seven coding sequences, mostly truncated.

Prokineticin domain-containing peptides

The sialotranscriptome of ticks has revealed proteins with the PFAM prokineticin domain. Prokineticins are peptides secreted by the suprachiasmatic nucleus of mammals and associated with circadian rhythm (http://pfam.sanger.ac.uk/family?acc=PF06607). Interestingly, a protein inducing smooth muscle contraction from the venom of the black mamba snake is a member of this family [141]. Four full-length CDS from this family were extracted from the sialotranscriptome of *Am. maculatum*, three of which are more than 20% divergent from each other and may derive from different genes. No function is known for any of the tick prokineticin domain salivary proteins.

Serum amyloid domain-containing proteins

In mammals, this protein family is a marker of acute-phase response and has been associated with endothelial function and tumor growth [142,143]. Members of this family have been found in previous sialotranscriptome of ticks, but none of their functions is known. Four full-length CDS have been retrieved from the sialotranscriptome of *Am. maculatum*, three of which appear to be alleles of the same gene.

Mucins

These are serine/threonine-rich proteins that can be heavily glycosylated with N-acetyl galactosamine residues [144,145]. They are commonly found in mucosal tissues, where they serve a role in mechanical protection of the cells and in preventing pathogen invasion. Their primary structure can be diverse; tick proteins of this family do not share significant sequence similarities to non-tick proteins except for the pattern of some repeated amino acids. Their galactosylation sites can be recognized by the site NetOGlyc [146]. Am-17147, constructed from 489 reads, has a signal peptide indicative of secretion and 58 glycosylation sites predicted by the NetOGlyc server. It shares only 29% identity to its *I. scapularis* homolog. Fifteen other putative mucins are indicated in Spreadsheet S2.

Secreted immunity-related products

Antimicrobial peptides of the defensin, hebrein/microplusin/Ricinusin, and lysozyme families [147] are represented in Spreadsheet S2, as well as peptides weakly similar to plant antimicrobials. Pathogen-recognition proteins of the peptidoglycan-recognition family, of the ixoderin/ficolin family of fibrinogen-domain containing proteins [148,149,150], and of the ML domain family [2,151] are also represented.

Tick-specific protein families

The sialotranscriptome of ticks abound with glycine-rich proteins, many of which were identified as members of the salivary cement used to attach the tick to its host's skin [152,153,154,155]. Members of the Ixodes genus also contain collagen-like glycine rich proteins. Additionally, cuticle and perithrophin proteins as well as some classes of antimicrobial peptides are rich in glycine and are included in this broad class [2]. We call attention to a large gene expansion indicated by the family

GRP 40-21 (Spreadsheet S2), where 23 CDS were extracted, including Am-2829 obtained from 1184 reads, which may represent the most abundant cement protein of *Am. maculatum.*

This protein is only 38% identical to an *Am. variegatum* homolog, indicating a fast evolution of this multigene family. Another group of proteins, both glycine and tyrosine rich, belongs to the
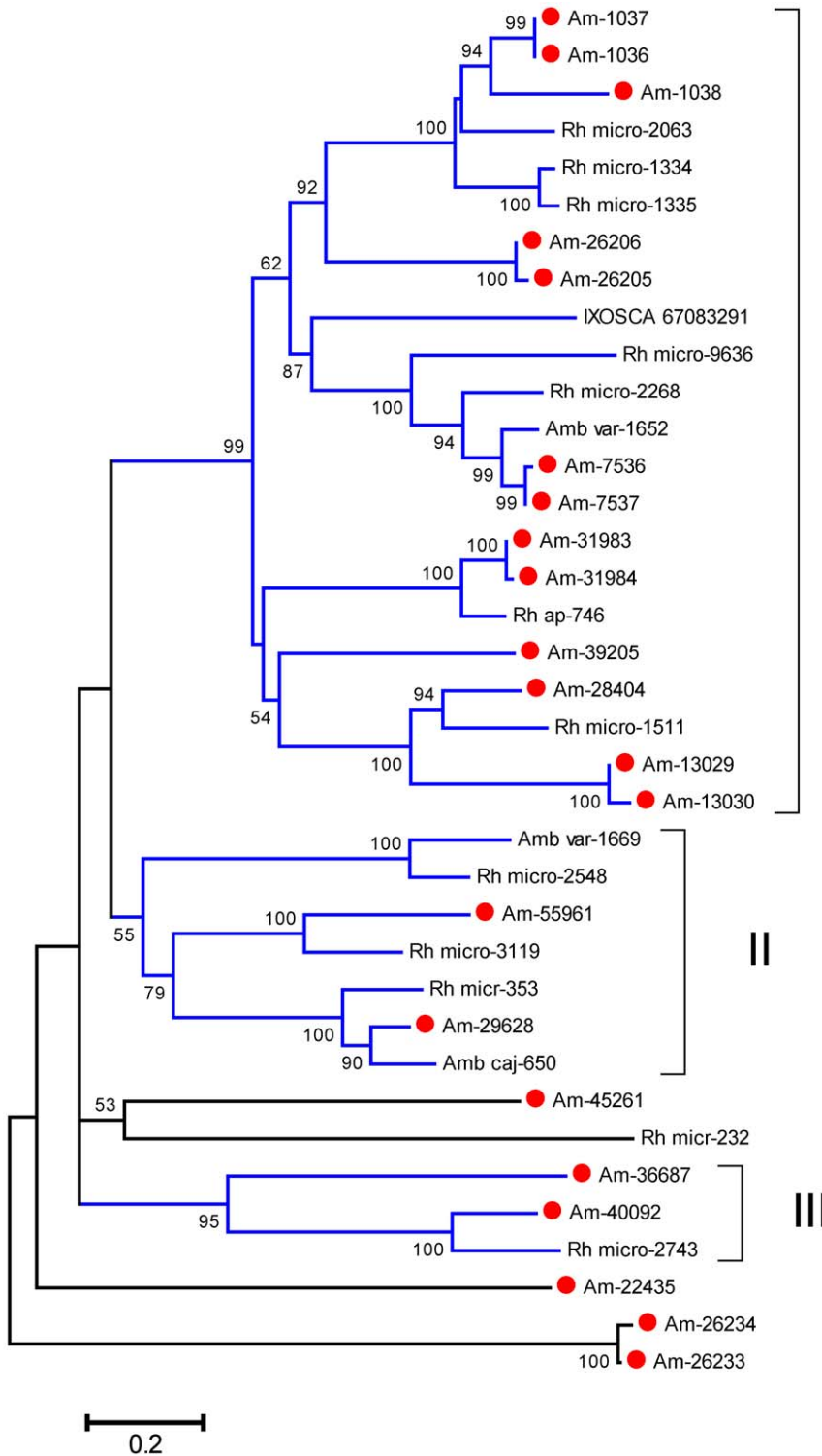

Figure 9. Phylogram of tick proteins from the novel 40-33 family. The bootstrapped phylogram (1,000 iterations) was obtained from the alignment of deducted *Amblyomma maculatum* proteins with homologs deducted from DBEST sequences described in a previous publication [2]. The *Am. maculatum* protein names start with Am- and are recognized by a red circle marker; the *Ixodes scapularis* protein is named IXOSCA followed by the first three letters of the species name followed by their NCBI gi| accession number. The number at the nodes indicates the bootstrap support above 50%, and the bar at the bottom indicates 20% amino acid divergence. Clades are indicated with Roman numerals.
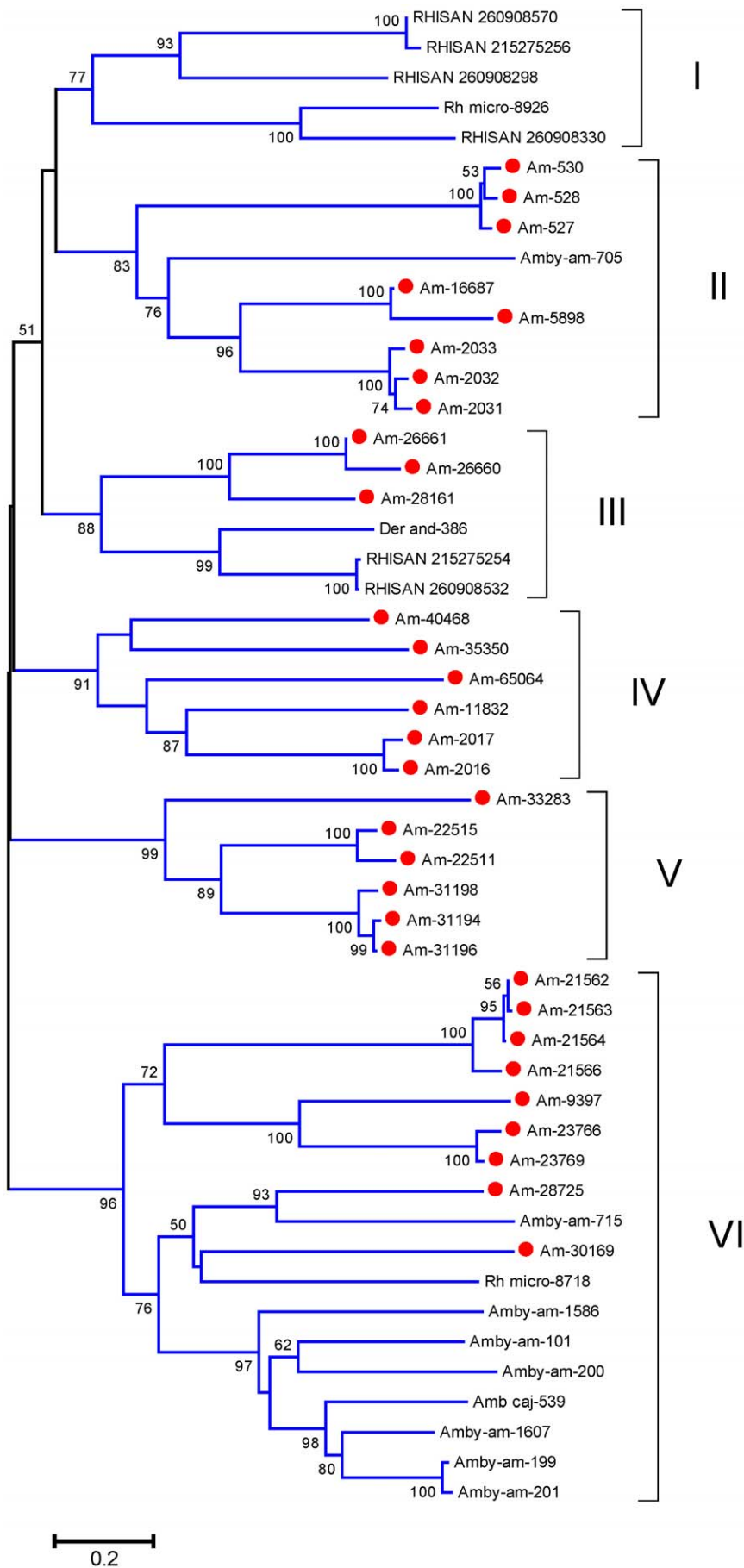doi:10.1371/journal.pone.0028525.g009

Figure 10. The evasin family of metastriate ticks. The bootstrapped phylogram (1,000 iterations) was obtained from the alignment of deducted *Amblyomma maculatum* proteins with homologs deduced from DBEST sequences described in a previous publication [2], and the *R. sanguineus* evasins from GenBank. The *Am. maculatum* protein names start with Am- and are recognized by a red circle marker. Other name conventions are as described in the previous figures. The number at the nodes indicates the bootstrap support above 50%, and the bar at the bottom indicates 20% amino acid divergence. Clades and superclades with strong bootstrap support are indicated with Roman numerals.
doi:10.1371/journal.pone.0028525.g010

previously named large GGY family [2]. Am-1905 is a highly expressed member of this family, being assembled from 1,932 reads.

Ixodegrins. This tick family, members of which range from 100–200 aa, was initially identified in *I. pacificus* [156] and found to be abundantly expressed in tick sialotranscriptomes [2]. A reverse position matrix made from the alignment of homologous proteins obtained from previous sialotranscriptomes was used by the BLAST package tool rpsblast to identify *Am. maculatum* members of this family, 20 of which are shown in Spreadsheet S2. Some of these proteins have weak matches to the PFAM prokineticin domain, as well as to the colipase domain, indicating Ixodegrins may be related to the prokineticin-domain proteins identified above. No member of this family has been functionally characterized thus far.

Disintegrins. This name derives from snake venom toxins having the RGD domain flanked by cysteines, which mimics the fibrinogen motif that binds to their platelet integrin receptor. Disintegrins thus occupy the integrin site destined for fibrinogen in the platelet surface, preventing platelet aggregation, because aggregation of platelets occurs by crosslinking of platelets by fibrinogen, which has multiple RGD sites [1,140,157]. Variations of the RGD domain exist, including KGD, KTS, and RTS domains [158]. The sialotranscriptome of *Am. maculatum* identified several members of this family, including five members of the related but divergent disintegrin 40–270 family (Figure 6A). Phylogenetic analysis is indicative of at least three genes coding for this family (Figure 6). Spreadsheet S2 displays additional polypeptides containing the RGD motif as well as the RTS motif.

Salp15. A salivary protein from *I. scapularis* was shown to inhibit CD4$^+$ T cell activation and to be important for Lyme's disease transmission [5,159,160,161,162]. A PFAM domain from this family (PF12115), constructed from *Ixodes* and *Argas* proteins, identified ten very divergent members of this family in *Am. maculatum* with relatively high E values. Members of this family have not been found in metastriate ticks previously. These ten *Am. maculatum* members can be divided into four families, containing four, two, two, and two proteins each, coding for proteins of mature size ranging from 10 to 26 kDa (Spreadsheet S2). The family with larger molecular weight (MW) has many galactosylation sites indicative of mucin.

Basic tail/18.3-kDa superfamily. This superfamily was identified in *I. scapularis* and then found to abound in other tick sialotranscriptomes. The "basic tail" name derives from a cluster of basic amino acids on the carboxyterminus of the protein. This basic tail may direct the protein to negatively charged phospholipids [163,164] that serve as scaffolding for the proteinase complexes that activate clotting [1]. The 18.3-kDa family does not have this basic tail, but was identified by using the tool PSI-BLAST searching the NR database starting with a basic tail protein. A PFAM domain named TSGP1 helps to identify both members of the family. Spreadsheet S2 identifies 40 members of this superfamily, either having the PFAM TSGP1 domain, or having similarities to members of the superfamily. Most CDS are full length and with a signal peptide indicative of secretion. Several of these sequences have indication of abundant galactosylation, indicating that they might serve as mucins. Only one member of

the family, named salp9, has been functionally characterized as an anticlotting protein [165].

23-kDa family. This protein family has, exclusive of soft and hard ticks [2], no known function. The sialotranscriptome of *Am. maculatum* identifies eight members of the family, as revealed by rpsblast using a model based on the previously known sequences (Additional spreadsheet S2). Phylogenetic analysis of these proteins resulting from their alignment with tick proteins found in the NR database by BLAST search with an E value smaller than 1e-20 (Figure 7) shows strong support for a clade containing several *I. scapularis* proteins, one *Ornithodoros* protein, and two very similar *Am. maculatum* proteins, possible alleles (Figure 7, Clade I). Additional clades indicate another six divergent genes coding for this protein family in *Am. maculatum*. Messages for Am-6910 and Am-4586 are abundant, the contigs having been assembled with over 400 reads.

8.9 kDa family. This is a protein family, exclusive of hard ticks, coding for secreted peptides of mature MW near 10 kDa, but also having proteins of double the size containing two domains. Their function is unknown. Coding sequences for 83 members of this family are reported for *Am. maculatum* in Spreadsheet S2. Some members of this family in *Am. maculatum* are abundantly expressed, their CDS being assembled with over 1,000 reads (Am-3961 and Am-733). Phylogenetic analysis of these proteins with their best matches from the NR database is complex, showing robust clades with both prostrate and metastriate sequences as well as clades showing expansion of *Am. maculatum* genes (not shown), suggestive of multiple rounds of gene duplication and fast evolution of this family.

Ixostatin. Ixostatins are an expanded protein family found in *I.s scapularis* and *I. pacificus* [12,156], but a few members were found in *Dermacentor* and *Amblyomma* [2]. Three members of this family are found in *Am. maculatum*, all poorly expressed, their CDS being assembled from 11–40 reads.

One-of-each family. This family was so named because only one protein per tick species has been found in sialotranscriptomes [2]; however, 24 CDS coding for members of this family were found in the sialotranscriptome of *Am. maculatum*.

PSI-BLAST search of the NR database using Am-18834 as query retrieves only tick proteins, indicating the specificity of the family (Additionalal file S3); however, several *I. scapularis* additional members of this family were found, indicating it to be a multigene family, as it appears to be in *Am. maculatum*. Alignment of members of the family having more than 200 aa indicates a very diverse family with few conserved amino acids (Figure 8A). The phylogram shows two robust superclades—one with *Amblyoma* and *Rhipicephalus* sequences and another with metastriate and prostrate proteins (Figure 8B)—and indicate at least 14 genes coding for this family in *Am. maculatum* (indicated by sequences >20% distant from each other) and 5 for *I. scapularis*.

Novel family 40-33. The sialotranscriptome of *Am. maculatum* reveals 21 CDS that are related at the 40% identity level, constituting the 33rd most abundant cluster at this level of protein clusterization. The predicted peptides have mature MW near 22 kDa. PSI-BLAST of Am-1037 against the NR database retrieves, after five iterations, spider and acari proteins, and at later iterations retrieves also *Daphnia* and insect proteins, indicating

this to be an unknown arthropod family (Additional file S4). Phylogenetic analysis of the alignment of the *Am. maculatum* members of this family with previously described tick proteins obtained from assembly of DBEST ESTs from ticks [2], as well as *I. scapularis* sequences deposited in GenBank, reveals several clades

containing sequences from multiple tick species (indicated by Roman numerals in Figure 9), including clade I with strong bootstrap support and containing both metastriate and prostriate proteins. Clades II and III have metastriate proteins only. There are possibly 14 genes coding for this family in *Am. maculatum*, based
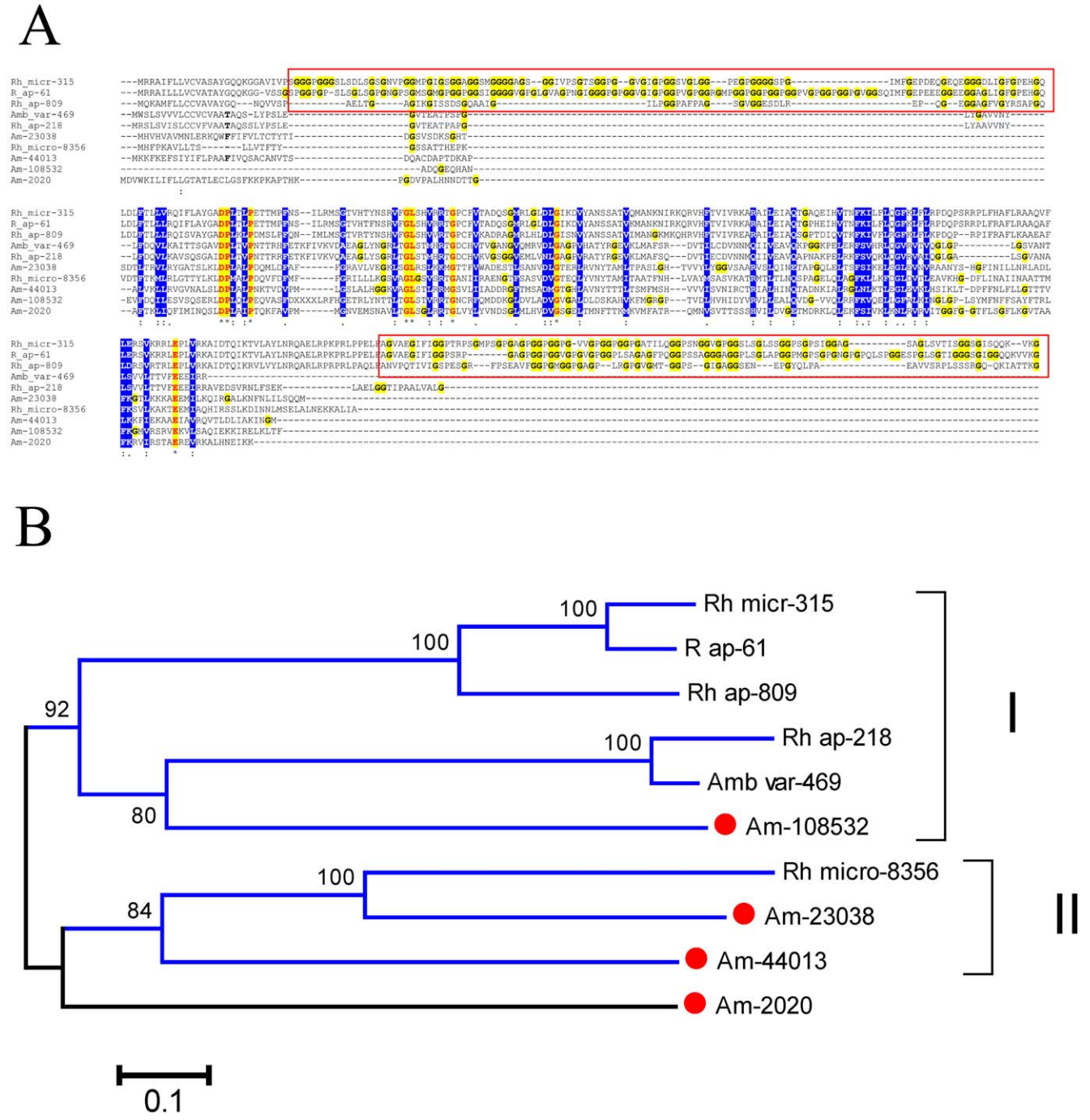

Figure 11. The salivary 40–279 novel family of metastriate ticks. (A) Clustal alignment. The symbols at the bottom indicate (*) identity of residues, (:) conserved and (.) less conserved residues. Identical residues are represented with red letters and yellow background. Conserved residues are shown in blue background. Glycines are shown in yellow background. The red boxes mark the glycine rich regions of 3 *Rhipicephalus* sequences. (B) Bootstraped phylogram of the alignment in A. The numbers at the nodes indicate the bootstrap support and the bar at the bottom indicates 10% amino acid divergence. The *Amblyomma maculatum* protein names start with Am- and are recognized by a red circle marker. Remaining sequences are from *Rhipicephalus microplus, R. appendiculatus* and *A. variegatum*. Clades with strong bootstrap support are indicated with Roman numerals. doi:10.1371/journal.pone.0028525.g011

on divergences larger than 20% amino acid sequence. Four members of this family in *Am. maculatum* were assembled with more than 300 reads, indicating relatively high expression levels.

Similar to Rhipicephalus and Ixodes proteins. Eight coding sequences from *Am. maculatum* sialotranscriptome present similarities with deducted protein sequences from *Rhipicephalus appendiculatus* deducted from ESTs deposited in DBEST, and a weaker match to an *I. scapularis* sequence deposited in GenBank. No similarities are found for other proteins in the NR database, thus characterizing a novel tick protein family.

Metastriate-specific protein families

Evasins. These are chemokine-binding proteins previously identified from the tick *R. sanguineus* [166,167]. Thirty-eight CDS from *Am. maculatum* were identified as similar to evasins based on similarity matches by BLAST or by rpsblast using evasin reverse-position matrices. Phylogenetic analysis (Figure 10) indicates the diversity of the family, even within the canonical *R. sanguineus* sequences, which are grouped in clades I and III. The *Am. maculatum* sequences are found in several clades, including in clade III. Analysis indicates at least 18 different genes coding for members of this family within *Am. maculatum*, based on sequences that are >20% divergent at the amino acid level. Relatively to other CDS from *Am. maculatum*, this family is not particularly highly expressed, the largest number of reads being 247 for Am-9397. If these proteins in *Am. maculatum* function as chemokine binders, they need not be at high concentrations, as these host proteins occur at nanomolar concentrations [121].

DAP-36 immunosuppressant. A salivary protein of 36 kDA was previously identified as an immunosuppressant from the tick *Dermacentor andersoni* [168,169] and found to be divergent and widespread in metastriate ticks [2]. Several members of this family are found in *Am. maculatum*, including two CDS that were assembled with more than 500 reads (Am-7601 and Am-5452). Sequence comparisons indicate at least eight genes coding for members of this family in *Am. maculatum* (data not shown).

Metastriate 13-kDa family. Fifteen related CDS coding for proteins with mature MW, mostly from 10–11 kDa, were found in the sialotranscriptome of *Am. maculatum* that match proteins previously classified under this name [2]. The function of these proteins is unknown.

Deorphanized metastriate-specific protein families

The sialotranscriptome of *Am. maculatum* allowed the identification of 12 clusters of sequences that match previously metastriate orphan proteins (orphan proteins being defined as those without primary sequence similarity when compared to the NR protein database of NCBI by blastp with the complexity filter turned off), thus characterizing 12 novel tick protein families. We will here analyze a single family, the metastriate novel family 40–279 (Spreadsheet S2), composed of four related CDS from *A. maculatum*. Alignment of the deducted protein sequences with their matches to a tick salivary protein database [2] shows that the proteins from *Rhipicephalus* have two glycine/proline-rich regions at either end of the sequence, and a middle part that has several sites of conservation among all the sequences (Figure 11A). The phylogram (Figure 11B) shows two clades with strong bootstrap support, indicating the existence of at least four genes for *Am. maculatum*, two for *R. microplus*, and three for *R. appendiculatus*. The conserved region allows the identification of a conserved amino acid block, [FL]-x(2)-[LVMI]-[VLI]-x(9)-D-P-

[LM]-x-[LVI]-P-x(18,24)-[VL]-x-G-L-x(2)-[VML]-x-[RK]-x-G-x(15)-[DN]-[LVM]-G-x(3)-[VL]-x(7)-[VLIM]-x(18,21)-[RQH]-x(2)-[LV]-x(3)-[QE]-x(6,7)-[VL]-x(2)-[FL]-x-[IVL]-x(2)-[LF]-x(4)-[LV]-x-[LVI]-x(12,21)-[LF]-x(3)-[VLI]-x(4)-E-x(2)-[VIL], that might help to identify these proteins. AM-2020 is well expressed, with 435 reads.

Amblyomma-specific families

Three hundred seven CDS, grouped in 54 families, provide significant matches solely to previously described *Amblyomma* proteins or to other proteins deducted from the *Am. maculatum* sialotranscriptome. Spreadsheet S2 contains 1,118 CDS coding for putative secreted proteins that are not classified. With additional tick genome and transcriptome sequencing, these proteins should be deorphanized.

Conclusions

This catalogue of salivary transcripts of *Amblyomma maculatum* herein reported represents a revolution since tick sialotranscriptomes were done 9 years ago for the first time [13,14]. The current assembly of over 1.5 million sequences from a normalized library, instead of thousands of sequences, allowed for a depth of transcript coverage thus far unequalled in sialotranscriptomes, and the deposition of 4,850 sequences to GenBank, as compared to very few in those pioneer publications. A total of 3,475 contigs were associated with a secretory function, and these were classified into 133 families. Twelve previously orphan metastriate salivary protein families were deorphanized with the current transcriptome; 55 families were identified as *Amblyomma* specific, until other transcriptomes may deorphanize them or confirm their unique status. This data set will serve as a platform for mining new pharmacologically active proteins and for development of anti tick vaccines to deter tick feeding or the pathogens they transmit.

Supporting Information

Table S1
(DOC)

Spreadsheet S1
(TXT)

Spreadsheet S2
(XLSX)

References

1. Francischetti IM (2010) Platelet aggregation inhibitors from hematophagous animals. Toxicon 56: 1130–1144.

2. Francischetti IMB, Sá-Nunes A, Mans BJ, Santos IM, Ribeiro JMC (2009) The role of saliva in tick feeding. Frontiers in Biosciences 14: 2051–2088.

3. Jones LD, Matthewson M, Nuttall PA (1992) Saliva-activated transmission (SAT) of Thogoto virus: dynamics of SAT factor activity in the salivary glands of *Rhipicephalus appendiculatus*, *Amblyomma variegatum*, and *Boophilus microplus* ticks. Exp Appl Acarol 13: 241–248.

4. Cavassani KA, Aliberti JC, Dias AR, Silva JS, Ferreira BR (2005) Tick saliva inhibits differentiation, maturation and function of murine bone-marrow-derived dendritic cells. Immunology 114: 235–245.

5. Ramamoorthi N, Narasimhan S, Pal U, Bao F, Yang XF, et al. (2005) The Lyme disease agent exploits a tick protein to infect the mammalian host. Nature 436: 573–577.

6. Kovar L (2004) Tick saliva in anti-tick immunity and pathogen transmission. Folia Microbiol (Praha) 49: 327–336.

7. Wikel SK (1999) Tick modulation of host immunity: an important factor in pathogen transmission. Int JParasitol 29: 851–859.

8. Titus RG, Bishop JV, Mejia JS (2006) The immunomodulatory factors of arthropod saliva and the potential for these factors to serve as vaccine targets to prevent pathogen transmission. Parasite Immunol 28: 131–141.

9. Nuttall PA, Trimnell AR, Kazimirova M, Labuda M (2006) Exposed and concealed antigens as vaccine targets for controlling ticks and tick-borne diseases. Parasite Immunol 28: 155–163.

10. Valenzuela JG (2004) Exploring tick saliva: from biochemistry to 'sialomes' and functional genomics. Parasitology 129 Suppl: S83–94.

11. Mans BJ, Neitz AW (2004) Adaptation of ticks to a blood-feeding environment: evolution from a functional perspective. Insect Biochem Mol Biol 34: 1–17.

12. Ribeiro JM, Alarcon-Chaidez F, Francischetti IM, Mans BJ, Mather TN, et al. (2006) An annotated catalog of salivary gland transcripts from *Ixodes scapularis* ticks. Insect Biochem Mol Biol 36: 111–129.

13. Valenzuela JG, Francischetti IMB, Pham VM, Garfield MK, Mather TN, et al. (2002) Exploring the sialome of the tick, *Ixodes scapularis*. J Exp Biol 205: 2843–2864.

14. Nene V, Lee D, Quackenbush J, Skilton R, Mwaura S, et al. (2002) AvGI, an index of genes transcribed in the salivary glands of the ixodid tick *Amblyomma variegatum*. Int J Parasitol 32: 1447–1456.

15. Ribeiro JM, Anderson JM, Manoukis NC, Meng Z, Francischetti IM (2011) A further insight into the sialome of the tropical bont tick, *Amblyomma variegatum*. BMC Genomics 12: 136.

16. Aljamali MN, Hern L, Kupfer D, Downard S, So S, et al. (2009) Transcriptome analysis of the salivary glands of the female tick *Amblyomma americanum* (Acari: Ixodidae). Insect Mol Biol 18: 129–154.

17. Batista IF, Chudzinski-Tavassi AM, Faria F, Simons SM, Barros-Batestti DM, et al. (2008) Expressed sequence tags (ESTs) from the salivary glands of the tick *Amblyomma cajennense* (Acari: Ixodidae). Toxicon 51: 823–834.

18. Alarcon-Chaidez FJ, Sun J, Wikel SK (2007) Transcriptome analysis of the salivary glands of *Dermacentor andersoni* Stiles (Acari: Ixodidae). Insect Biochem Mol Biol 37: 48–71.

19. Santos IK, Valenzuela JG, Ribeiro JM, de Castro M, Costa JN, et al. (2004) Gene discovery in *Boophilus microplus*, the cattle tick: the transcriptomes of ovaries, salivary glands, and hemocytes. Ann N Y Acad Sci 1026: 242–246.

20. Anatriello E, Ribeiro JM, de Miranda-Santos IK, Brandao LG, Anderson JM, et al. (2010) An insight into the sialotranscriptome of the brown dog tick, *Rhipicephalus sanguineus*. BMC Genomics 11: 450.

21. Francischetti IM, Mans BJ, Meng Z, Gudderra N, Veenstra TD, et al. (2008) An insight into the sialome of the soft tick, *Ornithodorus parkeri*. Insect Biochem Mol Biol 38: 1–21.

22. Francischetti IM, Meng Z, Mans BJ, Gudderra N, Hall M, et al. (2008) An insight into the salivary transcriptome and proteome of the soft tick and vector of epizootic bovine abortion, *Ornithodoros coriaceus*. J Proteomics 71: 493–512.

23. Chmelar J, Anderson JM, Mu J, Jochim RC, Valenzuela JG, et al. (2008) Insight into the sialome of the castor bean tick, *Ixodes ricinus*. BMC Genomics 9: 233.

24. Mans BJ, Andersen JF, Schwan TG, Ribeiro JM (2008) Characterization of anti-hemostatic factors in the argasid, *Argas monolakensis*: Implications for the evolution of blood-feeding in the soft tick family. Insect Biochem Mol Biol 38: 22–41.

25. Bishop FC, Hixson H (1936) Biology and economic importance of the Gulf Coast tick. J Econ Entomol 29: 1068–1076.

26. Espinoza-Gomez F, Newton-Sanchez O, Flores-Cazares G, De la Cruz-Ruiz M, Melnikov V, et al. (2011) Tick Paralysis Caused by *Amblyomma maculatum* on the Mexican Pacific Coast. Vector Borne Zoonotic Dis 11: 945–946.

27. Paffenbarger RS, Jr. (1951) Tick paralysis; implicating *Amblyomma maculatum*. New Orleans Med Surg J 103: 329–332.

28. Ewing SA, DuBois JG, Mathew JS, Panciera RJ (2002) Larval Gulf Coast ticks (*Amblyomma maculatum*) [Acari: Ixodidae] as host for *Hepatozoon americanum* [Apicomplexa: Adeleorina]. Vet Parasitol 103: 43–51.

29. Little SE, Allen KE, Johnson EM, Panciera RJ, Reichard MV, et al. (2009) New developments in canine hepatozoonosis in North America: a review. Parasit Vectors 2 Suppl 1: S5.

30. Mathew JS, Ewing SA, Panciera RJ, Kocan KM (1999) Sporogonic development of *Hepatozoon americanum* (Apicomplexa) in its definitive host, *Amblyomma maculatum* (Acarina). J Parasitol 85: 1023–1031.

31. Sumner JW, Durden LA, Goddard J, Stromdahl EY, Clark KL, et al. (2007) Gulf Coast ticks (*Amblyomma maculatum*) and *Rickettsia parkeri*, United States. Emerg Infect Dis 13: 751–753.

32. Patrick CD, Hair JA (1975) Laboratory rearing procedures and equipment for multi-host ticks (Acarina: Ixodidae). J Med Entomol 12: 389–390.

33. Meyer E, Aglyamova GV, Wang S, Buchanan-Carter J, Abrego D, et al. (2009) Sequencing and de novo analysis of a coral larval transcriptome using 454 GSFlx. BMC Genomics 10: 219.

34. Schwarz A, Helling S, Collin N, Teixeira CR, Medrano-Mercado N, et al. (2009) Immunogenic salivary proteins of *Triatoma infestans*: development of a recombinant antigen for the detection of low-level infestation of triatomines. PLoS Negl Trop Dis 3: e532.

35. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389–3402.

36. Huang X (1992) A contig assembly program based on sensitive detection of fragment overlaps. Genomics 14: 18–25.

37. Nielsen H, Engelbrecht J, Brunak S, von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10: 1–6.

38. Lewis S, Ashburner M, Reese MG (2000) Annotating eukaryote genomes. Curr Opin Struct Biol 10: 349–354.

39. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, et al. (2002) CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 30: 281–283.

40. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, et al. (2003) The COG database: an updated version includes eukaryotes. BMC Bioinformatics 4: 41.

41. Bateman A, Birney E, Durbin R, Eddy SR, Howe KL, et al. (2000) The Pfam protein families database. Nucleic Acids Res 28: 263–266.

42. Schultz J, Copley RR, Doerks T, Ponting CP, Bork P (2000) SMART: a web-based tool for the study of genetically mobile domains. Nucleic Acids Res 28: 231–234.

43. Sonnhammer EL, von Heijne G, Krogh A (1998) A hidden Markov model for predicting transmembrane helices in protein sequences. Proc Int Conf Intell Syst Mol Biol 6: 175–182.

44. Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, et al. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J 15: 115–130.

45. Duckert P, Brunak S, Blom N (2004) Prediction of proprotein convertase cleavage sites. Protein Eng Des Sel 17: 107–112.

46. Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Brief Bioinform 5: 150–163.

47. Jeanmougin F, Thompson JD, Gouy M, Higgins DG, Gibson TJ (1998) Multiple sequence alignment with Clustal X. Trends Biochem Sci 23: 403–405.

48. Pichu S, Yalcin EB, Ribeiro JM, King RS, Mather TN (2011) Molecular characterization of novel sulfotransferases from the tick, *Ixodes scapularis*. BMC Biochem 12: 32.

49. Yalcin EB, Stangl H, Pichu S, Mather TN, King RS (2010) Monoamine neurotransmitters as substrates for novel tick sulfotransferases, homology modeling, molecular docking, and enzyme kinetics. ACS Chem Biol 6: 176–184.

50. Kaufman W (1976) The influence of various factors on fluid secretion by in vitro salivary glands of ixodid ticks. J Exp Biol 64: 727–742.

51. Schmidt SP, et al. (1982) A dopamine sensitive adenylate cyclase in the salivary glands of *Amblyomma americanum* (L). Comp Biochem Physiol 72C: 8.

52. Ferrer-Sueta G, Radi R (2009) Chemical biology of peroxynitrite: kinetics, diffusion, and radicals. ACS Chem Biol 4: 161–177.

53. Rubbo H, Trostchansky A, O'Donnell VB (2009) Peroxynitrite-mediated lipid oxidation and nitration: mechanisms and consequences. Arch Biochem Biophys 484: 167–172.

54. Wilson AD, Heesom KJ, Mawby WJ, Mellor PS, Russell CL (2008) Identification of abundant proteins and potential allergens in *Culicoides nubeculosus* salivary glands. Vet Immunol Immunopathol 122: 94–103.

55. Campbell CL, Vandyke KA, Letchworth GJ, Drolet BS, Hanekamp T, et al. (2005) Midgut and salivary gland transcriptomes of the arbovirus vector *Culicoides sonorensis* (Diptera: Ceratopogonidae). Insect Mol Biol 14: 121–136.

56. Andersen JF, Pham VM, Meng Z, Champagne DE, Ribeiro JM (2009) Insight into the Sialome of the Black Fly, *Simulium vittatum*. J Proteome Res 8: 1474–1488.

57. Monteiro RQ, Rezaie AR, Bae JS, Calvo E, Andersen JF, et al. (2008) Ixolaris binding to factor X reveals a precursor state of factor Xa heparin-binding exosite. Protein Sci 17: 146–153.

58. Nazareth RA, Tomaz LS, Ortiz-Costa S, Atella GC, Ribeiro JM, et al. (2006) Antithrombotic properties of Ixolaris, a potent inhibitor of the extrinsic pathway of the coagulation cascade. Thromb Haemost 96: 7–13.

59. Monteiro RQ, Rezaie AR, Ribeiro JM, Francischetti IM (2005) Ixolaris: a factor Xa heparin-binding exosite inhibitor. Biochem J 387: 871–877.

60. Francischetti IM, Mather TN, Ribeiro JM (2004) Penthalaris, a novel recombinant five-Kunitz tissue factor pathway inhibitor (TFPI) from the salivary gland of the tick vector of Lyme disease, *Ixodes scapularis*. Thromb Haemost 91: 886–898.

61. Calvo E, Sanchez-Vargas I, Kotsyfakis M, Favreau AJ, Barbian KD, et al. (2010) The salivary gland transcriptome of the eastern tree hole mosquito, *Ochlerotatus triseriatus*. J Med Entomol 47: 376–386.

62. Lucchesi KJ, Moczydlowski E (1991) On the interaction of bovine pancreatic trypsin inhibitor with maxi Ca(2+)-activated K+ channels. A model system for analysis of peptide-induced subconductance states. J Gen Physiol 97: 1295–1319.

63. Kwong PD, McDonald NQ, Sigler PB, Hendrickson WA (1995) Structure of beta 2-bungarotoxin: potassium channel binding by Kunitz modules and targeted phospholipase action. Structure 3: 1109–1119.

64. Harvey AL (1997) Recent studies on dendrotoxins and potassium ion channels. Gen Pharmacol 28: 7–12.

65. Castaneda O, Harvey AL (2009) Discovery and characterization of cnidarian peptide toxins that affect neuronal potassium ion channels. Toxicon 54: 1119–1124.

66. Paesen GC, Siebold C, Dallas ML, Peers C, Harlos K, et al. (2009) An ion-channel modulator from the saliva of the brown ear tick has a highly modified Kunitz/BPTI structure. J Mol Biol 389: 734–747.

67. Mans BJ, Louw AI, Neitz AW (2002) Evolution of hematophagy in ticks: common origins for blood coagulation and platelet aggregation inhibitors from soft ticks of the genus *Ornithodoros*. Mol Biol Evol 19: 1695–1705.

68. Li WF, Ma GX, Zhou XX (2006) Apidaecin-type peptides: biodiversity, structure-function relationships and mode of action. Peptides 27: 2350–2359.

69. Fogaca AC, Almeida IC, Eberlin MN, Tanaka AS, Bulet P, et al. (2005) Ixodidin, a novel antimicrobial peptide from the hemocytes of the cattle tick *Boophilus microplus* with inhibitory activity against serine proteinases. Peptides 27: 667–674.

70. Sasaki SD, de Lima CA, Lovato DV, Juliano MA, Torquato RJ, et al. (2008) BmSI-7, a novel subtilisin inhibitor from *Boophilus microplus*, with activity toward Pr1 proteases from the fungus *Metarhizium anisopliae*. Exp Parasitol 118: 214–220.

71. Lenarcic B, Ritonja A, Strukelj B, Turk B, Turk V (1997) Equistatin, a new inhibitor of cysteine proteinases from *Actinia equina*, is structurally related to thyroglobulin type-1 domain. J Biol Chem 272: 13899–13903.

72. Lenarcic B, Turk V (1999) Thyroglobulin type-1 domains in equistatin inhibit both papain-like cysteine proteinases and cathepsin D. J Biol Chem 274: 563–566.

73. Galesa K, Pain R, Jongsma MA, Turk V, Lenarcic B (2003) Structural characterization of thyroglobulin type-1 domains of equistatin. FEBS Lett 539: 120–124.

74. Kotsyfakis M, Sa-Nunes A, Francischetti IM, Mather TN, Andersen JF, et al. (2006) Antiinflammatory and immunosuppressive activity of sialostatin L, a salivary cystatin from the tick *Ixodes scapularis*. J Biol Chem 281: 26298–26307.

75. Kotsyfakis M, Anderson JM, Andersen JF, Calvo E, Francischetti IM, et al. (2008) Cutting edge: Immunity against a "silent" salivary antigen of the Lyme vector *Ixodes scapularis* impairs its ability to feed. J Immunol 181: 5209–5212.

76. Leboulle G, Crippa M, Decrem Y, Mejri N, Brossard M, et al. (2002) Characterization of a novel salivary immunosuppressive protein from *Ixodes ricinus* ticks. J Biol Chem 15: 15.

77. Prevot PP, Couvreur B, Denis V, Brossard M, Vanhamme L, et al. (2007) Protective immunity against *Ixodes ricinus* induced by a salivary serpin. Vaccine 25: 3284–3292.

78. Chmelar J, Oliveira CJ, Rezacova P, Francischetti IM, Kovarova Z, et al. (2011) A tick salivary protein targets cathepsin G and chymase and inhibits host inflammation and platelet aggregation. Blood 117: 736–744.

79. Muleng A, Sugino M, Nakajima M, Sugimoto C, Onuma M (2001) Tick-Encoded serine proteinase inhibitors (serpins); potential target antigens for tick vaccine development. J Vet Med Sci 63: 1063–1069.

80. Imamura S, da Silva Vaz Junior I, Sugino M, Ohashi K, Onuma M (2005) A serine protease inhibitor (serpin) from *Haemaphysalis longicornis* as an anti-tick vaccine. Vaccine 23: 1301–1311.

81. Rawlings ND, Tolle DP, Barrett AJ (2004) Evolutionary families of peptidase inhibitors. Biochem J 378: 705–716.

82. Kanost MR (1999) Serine proteinase inhibitors in arthropod immunity. Dev Comp Immunol 23: 291–301.

83. Watanabe RM, Soares TS, Morais-Zani K, Tanaka-Azevedo AM, Maciel C, et al. (2010) A novel trypsin Kazal-type inhibitor from *Aedes aegypti* with thrombin coagulant inhibitory activity. Biochimie 92: 933–939.

84. Zhou J, Liao M, Hatta T, Tanaka M, Xuan X, et al. (2006) Identification of a follistatin-related protein from the tick *Haemaphysalis longicornis* and its effect on tick oviposition. Gene 372: 191–198.

85. Li KC, Zhang FX, Li CL, Wang F, Yu MY, et al. (2011) Follistatin-like 1 suppresses sensory afferent transmission by activating Na+,K+-ATPase. Neuron 69: 974–987.

86. Arolas JL, Lorenzo J, Rovira A, Castella J, Aviles FX, et al. (2005) A carboxypeptidase inhibitor from the tick *Rhipicephalus bursa*: isolation, cDNA cloning, recombinant expression, and characterization. J Biol Chem 280: 3441–3448.

87. Arolas JL, Popowicz GM, Lorenzo J, Sommerhoff CP, Huber R, et al. (2005) The three-dimensional structures of tick carboxypeptidase inhibitor in complex with A/B carboxypeptidases reveal a novel double-headed binding mode. J Mol Biol 350: 489–498.

88. Gong H, Zhou J, Liao M, Hatta T, Harnnoi T, et al. (2007) Characterization of a carboxypeptidase inhibitor from the tick *Haemaphysalis longicornis*. J Insect Physiol 53: 1079–1087.

89. Hengst U, Albrecht H, Hess D, Monard D (2001) The phosphatidylethanol-amine-binding protein is the prototype of a novel family of serine protease inhibitors. J Biol Chem 276: 535–540.

90. Mima J, Narita Y, Chiba H, Hayashi R (2003) The multiple site binding of carboxypeptidase Y inhibitor (IC) to the cognate proteinase. Implications for the biological roles of the phosphatidylethanolamine-binding protein. J Biol Chem 278: 29792–29798.

91. Mans BJ, Andersen JF, Francischetti IM, Valenzuela JG, Schwan TG, et al. (2008) Comparative sialomics between hard and soft ticks: Implications for the evolution of blood-feeding behavior. Insect Biochem Mol Biol 38: 42–58.

92. Harnnoi T, Sakaguchi T, Nishikawa Y, Xuan X, Fujisaki K (2007) Molecular characterization and comparative study of 6 salivary gland metalloproteases from the hard tick, *Haemaphysalis longicornis*. Comp Biochem Physiol B Biochem Mol Biol 147: 93–101.

93. Francischetti IM, Mather TN, Ribeiro JM (2003) Cloning of a salivary gland metalloprotease and characterization of gelatinase and fibrin(ogen)lytic activities in the saliva of the Lyme disease tick vector *Ixodes scapularis*. Biochem Biophys Res Commun 305: 869–875.

94. Francischetti IM, Mather TN, Ribeiro JM (2005) Tick saliva is a potent inhibitor of endothelial cell proliferation and angiogenesis. Thromb Haemost 94: 167–174.

95. Yang XF, Chatellard C, Lazure C, Crine P, Boileau G (1994) Insertion of hydrophilic amino acid residues in the signal peptide/membrane anchor domain of neprilysin (neutral endopeptidase-24.11) results in its cleavage: role of the position of insertion. Arch Biochem Biophys 315: 382–386.

96. Turner AJ, Brown CD, Carson JA, Barnes K (2000) The neprilysin family in health and disease. Adv Exp Med Biol 477: 229–240.

97. Nielsen H, Brunak S, von Heijne G (1999) Machine learning approaches for the prediction of signal peptides and other protein sorting signals. Protein Eng 12: 3–9.

98. Ribeiro JM, Mather TN (1998) *Ixodes scapularis*: salivary kininase activity is a metallo dipeptidyl carboxypeptidase. Exp Parasitol 89: 213–221.

99. Xu X, Yang H, Ma D, Wu J, Wang Y, et al. (2008) Toward an understanding of the molecular mechanism for successful blood feeding by coupling proteomics analysis with pharmacological testing of horsefly salivary glands. Mol Cell Proteomics 7: 582–590.

100. Alim MA, Tsuji N, Miyoshi T, Islam MK, Hatta T, et al. (2009) Legumains from the hard tick *Haemaphysalis longicornis* play modulatory roles in blood feeding and gut cellular remodelling and impact on embryogenesis. Int J Parasitol 39: 97–107.

101. Sojka D, Francischetti IM, Calvo E, Kotsyfakis M (2007) Cysteine proteases from bloodfeeding arthropod ectoparasites. Adv Exp Med Biol 712: 177–191.

102. Abdul Alim M, Tsuji N, Miyoshi T, Khyrul Islam M, Huang X, et al. (2007) Characterization of asparaginyl endopeptidase, legumain induced by blood feeding in the ixodid tick *Haemaphysalis longicornis*. Insect Biochem Mol Biol 37: 911–922.

103. Alim MA, Tsuji N, Miyoshi T, Islam MK, Hatta T, et al. (2008) Developmental stage- and organ-specific expression profiles of asparaginyl endopeptidases/legumains in the ixodid tick *Haemaphysalis longicornis*. J Vet Med Sci 70: 1363–1366.

104. Alim MA, Tsuji N, Miyoshi T, Islam MK, Huang X, et al. (2008) HlLgm2, a member of asparaginyl endopeptidases/legumains in the midgut of the ixodid tick *Haemaphysalis longicornis*, is involved in blood-meal digestion. J Insect Physiol 54: 573–585.

105. Horn M, Nussbaumerova M, Sanda M, Kovarova Z, Srba J, et al. (2009) Hemoglobin digestion in blood-feeding ticks: mapping a multipeptidase pathway by functional proteomics. Chem Biol 16: 1053–1063.

106. Sojka D, Hajdusek O, Dvorak J, Sajid M, Franta Z, et al. (2007) IrAE: an asparaginyl endopeptidase (legumain) in the gut of the hard tick *Ixodes ricinus*. Int J Parasitol 37: 713–724.

107. Ribeiro JM, Mans BJ, Arca B (2010) An insight into the sialome of blood-feeding Nematocera. Insect Biochem Mol Biol 40: 767–784.

108. Calvo E, Ribeiro JM (2006) A novel secreted endonuclease from *Culex quinquefasciatus* salivary glands. J Exp Biol 209: 2651–2659.

109. Wartha F, Beiter K, Normark S, Henriques-Normark B (2007) Neutrophil extracellular traps: casting the NET over pathogenesis. Curr Opin Microbiol 10: 52–56.

110. Sun D, McNicol A, James AA, Peng Z (2006) Expression of functional recombinant mosquito salivary apyrase: A potential therapeutic platelet aggregation inhibitor. Platelets 17: 178–184.

111. Faudry E, Lozzi SP, Santana JM, D'Souza-Ault M, Kieffer S, et al. (2004) *Triatoma infestans* apyrases belong to the 5′-nucleotidase family. J Biol Chem 279: 19607–19613.

112. Champagne DE, Smartt CT, Ribeiro JM, James AA (1995) The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5′-nucleotidase family. Proc Natl Acad Sci U S A 92: 694–698.

113. Stutzer C, Mans BJ, Gaspar AR, Neitz AW, Maritz-Olivier C (2009) *Ornithodoros savignyi*: soft tick apyrase belongs to the 5′-nucleotidase family. Exp Parasitol 122: 318–327.

114. Ogata S, Hayashi Y, Misumi Y, Ikehara Y (1990) Membrane-anchoring domain of rat liver 5′-nucleotidase: identification of the COOH-terminal serine-523 covalently attached with a glycolipid. Biochemistry 29: 7923–7927.

115. Misumi Y, Ogata S, Ohkubo K, Hirose S, Ikehara Y (1990) Primary structure of human placental 5'-nucleotidase and identification of the glycolipid anchor in the mature form. Eur J Biochem 191: 563–569.

116. Poisson G, Chauve C, Chen X, Bergeron A (2007) FragAnchor: a large-scale predictor of glycosylphosphatidylinositol anchors in eukaryote protein sequences by qualitative scoring. Genomics Proteomics Bioinformatics 5: 121–130.

117. Cerna P, Mikes L, Volf P (2002) Salivary gland hyaluronidase in various species of phlebotomine sand flies (Diptera: psychodidae). Insect Biochem Mol Biol 32: 1691–1697.

118. Volfova V, Hostomska J, Cerny M, Votypka J, Volf P (2008) Hyaluronidase of bloodsucking insects and its enhancing effect on *Leishmania* infection in mice. PLoS Negl Trop Dis 2: e294.

119. Ribeiro JM, Charlab R, Rowton ED, Cupp EW (2000) *Simulium vittatum* (Diptera: Simuliidae) and *Lutzomyia longipalpis* (Diptera: Psychodidae) salivary gland hyaluronidase activity. J Med Entomol 37: 743–747.

120. Neitz AW, Howell CJ, Potgieter DJ, Bezuidenhout JD (1978) Proteins and free amino acids in the salivary secretion and haemolymph of the tick *Amblyomma hebraeum*. Onderstepoort J Vet Res 45: 235–240.

121. Haessler U, Pisano M, Wu M, Swartz MA (2011) Dendritic cell chemotaxis in 3D under defined chemokine gradients reveals differential response to ligands CCL21 and CCL19. Proc Natl Acad Sci U S A 108: 5614–5619.

122. Laguri C, Arenzana-Seisdedos F, Lortat-Jacob H (2008) Relationships between glycosaminoglycan and receptor binding sites in chemokines-the CXCL12 example. Carbohydr Res 343: 2018–2023.

123. Nunn MA, Sharma A, Paesen GC, Adamson S, Lissina O, et al. (2005) Complement inhibitor of C5 activation from the soft tick *Ornithodoros moubata*. J Immunol 174: 2084–2091.

124. Ribeiro JMC, Arca B (2009) From sialomes to the sialoverse: An insight into the salivary potion of blood feeding insects. Adv Insect Physiol 37: 59–118.

125. Mans BJ, Ribeiro JM (2008) Function, mechanism and evolution of the moubatin-clade of soft tick lipocalins. Insect Biochem Mol Biol 38: 841–852.

126. Mans BJ, Ribeiro JM (2008) A novel clade of cysteinyl leukotriene scavengers in soft ticks. Insect Biochem Mol Biol 38: 862–870.

127. Mans BJ, Ribeiro JM, Andersen JF (2008) Structure, function, and evolution of biogenic amine-binding proteins in soft ticks. J Biol Chem 283: 18721–18733.

128. Sangamnatdej S, Paesen GC, Slovak M, Nuttall PA (2002) A high affinity serotonin- and histamine-binding lipocalin from tick saliva. Insect Mol Biol 11: 79–86.

129. Paesen GC, Adams PL, Nuttall PA, Stuart DL (2000) Tick histamine-binding proteins: lipocalins with a second binding cavity. Biochim Biophys Acta 1482: 92–101.

130. Paesen GC, Adams PL, Harlos K, Nuttall PA, Stuart DI (1999) Tick histamine-binding proteins: isolation, cloning, and three- dimensional structure. Mol Cell 3: 661–671.

131. Hoffman DR (1993) Allergens in Hymenoptera venom. XXV: The amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. J Allergy Clin Immunol 92: 707–716.

132. Yamazaki Y, Morita T (2004) Structure and function of snake venom cysteine-rich secretory proteins. Toxicon 44: 227–231.

133. Udby L, Sorensen OE, Pass J, Johnsen AH, Behrendt N, et al. (2004) Cysteine-rich secretory protein 3 is a ligand of alpha1B-glycoprotein in human plasma. Biochemistry 43: 12877–12886.

134. Yamazaki Y, Hyodo F, Morita T (2003) Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. Arch Biochem Biophys 412: 133–141.

135. Gibbs GM, Roelants K, O'Bryan MK (2008) The CAP superfamily: cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins–roles in reproduction, cancer, and immune defense. Endocr Rev 29: 865–897.

136. Wang X, Ribeiro JM, Broce AB, Wilkerson MJ, Kanost MR (2009) An insight into the transcriptome and proteome of the salivary gland of the stable fly, *Stomoxys calcitrans*. Insect Biochem Mol Biol 39: 607–614.

137. Ameri M, Wang X, Wilkerson MJ, Kanost MR, Broce AB (2008) An immunoglobulin binding protein (antigen 5) of the stable fly (Diptera: Muscidae) salivary gland stimulates bovine immune responses. J Med Entomol 45: 94–101.

138. Ma D, Wang Y, Yang H, Wu J, An S, et al. (2009) Anti-thrombosis repertoire of blood-feeding horsefly salivary glands. Mol Cell Proteomics 8: 2071–2079.

139. Ma D, Xu X, An S, Liu H, Yang X, et al. A novel family of RGD-containing disintegrins (Tablysin-15) from the salivary gland of the horsefly *Tabanus yao* targets alphaIIbbeta3 or alphaVbeta3 and inhibits platelet aggregation and angiogenesis. Thromb Haemost 105: 1032–1045.

140. Niewiraowski S, McLane MA, Kloczewiak M, Stewart GJ (1994) Disintegrins and other naturally occurring antagonists of platelet fibrinogen receptors. Sem Hematol 31: 289–300.

141. Boisbouvier J, Albrand JP, Blackledge M, Jaquinod M, Schweitz H, et al. (1998) A structural homologue of colipase in black mamba venom revealed by NMR floating disulphide bridge analysis. J Mol Biol 283: 205–219.

142. Kisilevsky R, Tam SP, Ancsin JB (2008) The anti-atherogenic potential of serum amyloid A peptides. Curr Opin Investig Drugs 9: 265–273.

143. Uhlar CM, Whitehead AS (1999) Serum amyloid A, the major vertebrate acute-phase reactant. Eur J Biochem 265: 501–523.

144. Brockhausen I (2006) Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. EMBO Rep 7: 599–604.

145. Voynow JA, Gendler SJ, Rose MC (2006) Regulation of mucin genes in chronic inflammatory airway diseases. Am J Respir Cell Mol Biol 34: 661–665.

146. Julenius K, Molgaard A, Gupta R, Brunak S (2005) Prediction, conservation analysis, and structural characterization of mammalian mucin-type O-glycosylation sites. Glycobiology 15: 153–164.

147. Kopacek P, Hajdusek O, Buresova V, Daffre S (2010) Tick innate immunity. Adv Exp Med Biol 708: 137–162.

148. Rego RO, Hajdusek O, Kovar V, Kopacek P, Grubhoffer L, et al. (2005) Molecular cloning and comparative analysis of fibrinogen-related proteins from the soft tick *Ornithodoros moubata* and the hard tick *Ixodes ricinus*. Insect Biochem Mol Biol 35: 991–1004.

149. Sterba J, Dupejova J, Fiser M, Vancova M, Grubhoffer L (2011) Fibrinogen-related proteins in ixodid ticks. Parasit Vectors 4: 127.

150. Rego RO, Kovar V, Kopacek P, Weise C, Man P, et al. (2006) The tick plasma lectin, Dorin M, is a fibrinogen-related molecule. Insect Biochem Mol Biol 36: 291–299.

151. Horackova J, Rudenko N, Golovchenko M, Havlikova S, Grubhoffer L (2010) IrML - a gene encoding a new member of the ML protein family from the hard tick, *Ixodes ricinus*. J Vector Ecol 35: 410–418.

152. Trimnell AR, Davies GM, Lissina O, Hails RS, Nuttall PA (2005) A cross-reactive tick cement antigen is a candidate broad-spectrum tick vaccine. Vaccine 23: 4329–4341.

153. Bishop R, Lambson B, Wells C, Pandit P, Osaso J, et al. (2002) A cement protein of the tick *Rhipicephalus appendiculatus*, located in the secretory e cell granules of the type III salivary gland acini, induces strong antibody responses in cattle. Int J Parasitol 32: 833–842.

154. Brown SJ, Shapiro SZ, Askenase PW (1984) Characterization of tick antigens inducing host immune resistance. I. Immunization of guinea pigs with *Amblyomma americanum*-derived salivary gland extracts and identification of an important salivary gland protein antigen with guinea pig anti-tick antibodies. J Immunol 133: 3319–3325.

155. Chinery WA (1973) The nature and origin of the "cement" substance at the site of attachment and feeding of adult *Haemaphysalis spinigera* (Ixodidae). J Med Entomol 10: 355–362.

156. Francischetti IM, My Pham V, Mans BJ, Andersen JF, Mather TN, et al. (2005) The transcriptome of the salivary glands of the female western black-legged tick *Ixodes pacificus* (Acari: Ixodidae). Insect Biochem Mol Biol 35: 1142–1161.

157. Huang TF, Niewriarowski S (1994) Disintegrins: The naturally-occurring antagonists of platelet fibrinogen receptors. J Toxicol Toxin Rev 13: 253–273.

158. Calvete JJ, Marcinkiewicz C, Sanz L (2007) KTS and RTS-disintegrins: anti-angiogenic viper venom peptides specifically targeting the alpha 1 beta 1 integrin. Curr Pharm Des 13: 2853–2859.

159. Juncadella IJ, Garg R, Ananthnarayanan SK, Yengo CM, Anguita J (2007) T-cell signaling pathways inhibited by the tick saliva immunosuppressor, Salp15. FEMS Immunol Med Microbiol 49: 433–438.

160. Garg R, Juncadella IJ, Ramamoorthi N, Ashish, Ananthanarayanan SK, et al. (2006) Cutting edge: CD4 is the receptor for the tick saliva immunosuppressor, Salp15. J Immunol 177: 6579–6583.

161. Anguita J, Ramamoorthi N, Hovius JW, Das S, Thomas V, et al. (2002) Salp15, an *Ixodes scapularis* salivary protein, inhibits CD4(+) T cell activation. Immunity 16: 849–859.

162. Hojgaard A, Biketov SF, Shtannikov AV, Zeidner NS, Piesman J (2009) Molecular identification of Salp15, a key salivary gland protein in the transmission of lyme disease spirochetes, from *Ixodes persulcatus* and *Ixodes pacificus* (Acari: Ixodidae). J Med Entomol 46: 1458–1463.

163. Andersen JF, Gudderra NP, Francischetti IM, Valenzuela JG, Ribeiro JM (2004) Recognition of anionic phospholipid membranes by an antihemostatic protein from a blood-feeding insect. Biochemistry 43: 6987–6994.

164. Stevenson KJ, Poller L (1982) The procoagulant activity of partial thromboplastin extracts: the role of phosphatidyl serine. Thromb Res 26: 341–350.

165. Narasimhan S, Koski RA, Beaulieu B, Anderson JF, Ramamoorthi N, et al. (2002) A novel family of anticoagulants from the saliva of *Ixodes scapularis*. Insect Mol Biol 11: 641–650.

166. Deruaz M, Frauenschuh A, Alessandri AL, Dias JM, Coelho FM, et al. (2008) Ticks produce highly selective chemokine binding proteins with antiinflammatory activity. J Exp Med 205: 2019–2031.

167. Frauenschuh A, Power CA, Deruaz M, Ferreira BR, Silva JS, et al. (2007) Molecular cloning and characterization of a highly selective chemokine-binding protein from the tick *Rhipicephalus sanguineus*. J Biol Chem 282: 27250–27258.

168. Bergman DK, Palmer MJ, Caimano MJ, Radolf JD, Wikel SK (2000) Isolation and molecular cloning of a secreted immunosuppressant protein from *Dermacentor andersoni* salivary gland. J Parasitol 86: 516–525.

169. Bergman DK, Ramachandra RN, Wikel SK (1998) Characterization of an immunosuppressant protein from *Dermacentor andersoni* (Acari: Ixodidae) salivary glands. J Med Entomol 35: 505–509.